A GEODATABASE OF HISTORICAL AND CONTEMPORARY OCEANOGRAPHIC DATASETS FOR INVESTIGATING THE ROLE OF THE PHYSICAL ENVIRONMENT IN SHAPING PATTERNS OF SEABED BIODIVERSITY IN THE GULF OF MAINE

M.E. Greenlaw, J.A. Sameoto, P. Lawton, N.H. Wolff, L.S. Incze, C.R. Pitcher, S.J. Smith and A. Drozdowski

Science Branch, Maritimes Region
Fisheries and Oceans Canada
Biological Station
531 Brandy Cove Road, St. Andrews, NB
E5B 2L9

2010

Canadian Technical Report of Fisheries and Aquatic Sciences 2895

Fisheries and Oceans Canada
PêchesetOcéans Canada
Canada

Canadian technical report of fisheries and aquatic sciences

Technical reports contain scientific and technical information that contributes to existing knowledge but which is not normally appropriate for primary literature. Technical reports are directed primarily toward a worldwide audience and have an international distribution. No restriction is placed on subject matter and the series reflects the broad interests and policies of the Department of Fisheries and Oceans, namely, fisheries and aquatic sciences.

Technical reports may be cited as full publications. The correct citation appears above the abstract of each report. Each report is abstracted in *Aquatic Sciences and Fisheries Abstracts* and indexed in the Department's annual index to scientific and technical publications.

Numbers 1-456 in this series were issued as Technical Reports of the Fisheries Research Board of Canada. Numbers 457-714 were issued as Department of the Environment, Fisheries and Marine Service, Research and Development Directorate Technical Reports. Numbers 715–924 were issued as Department of Fisheries and the Environment, Fisheries and Marine Service Technical Reports. The current series name was changed with report number 925.

Technical reports are produced regionally but are numbered nationally. Requests for individual reports will be filled by the issuing establishment listed on the front cover and title page. Out-of-stock reports will be supplied for a fee by commercial agents.

Rapport technique canadien des sciences halieutiques et aquatiques

Les rapports techniques contiennent des renseignements scientifiques et techniques qui constituent une contribution aux connaissances actuelles, mais qui ne sont pas normalement appropriés pour la publication dans un journal scientifique. Les rapports techniques sont destinés essentiellement à un public international et ils sont distribués à cet échelon. Il n'y a aucune restriction quant au sujet; de fait, la série reflète la vaste gamme des intérêts et des politiques du ministère des Pêches et des Océans, c'est-à-dire les sciences halieutiques et aquatiques.

Les rapports techniques peuvent être cités comme des publications complètes. Le titre exact parait au-dessus du résumé de chaque rapport. Les rapports techniques sont résumés dans la revue *Résumés des sciences aquatiques et halieutiques*, et ils sont classés dans l'index annuel des publications scientifiques et techniques du Ministère.

Les numéros 1 à 456 de cette série ont été publiés à titre de rapports techniques de l'Office des recherches sur les pêcheries du Canada. Les numéros 457 à 714 sont parus à titre de rapports techniques de la Direction générale de la recherche et du développement, Service des pêches et de la mer, ministère de l'Environnement. Les numéros 715 à 924 ont été publiés à titre de rapports techniques du Service des pêches et de la mer, ministère des Pêches et de l'Environnement. Le nom actuel de la série a été établi lors de la parution du numéro 925.

Les rapports techniques sont produits à l'échelon régional, mais numéroté à l'échelon national. Les demandes de rapports seront satisfaites par l'établissement auteur dont le nom figure sur la couverture et la page du titre. Les rapports épuisés seront fournis contre rétribution par des agents commerciaux.

Canadian Technical Report of
Fisheries and Aquatic Sciences 2895

2010

A Geodatabaseof Historical and Contemporary Oceanographic Datasets for Investigating the Role of the Physical Environment in Shaping Patterns of Seabed Biodiversity in the Gulf of Maine

by

M.E. Greenlaw, J.A. Sameoto[1], P. Lawton,N.H. Wolff[2], L.S. Incze[2],
C.R. Pitcher[3], S.J. Smith[1] and A. Drozdowski[1]

Fisheries and Oceans Canada, Science Branch, Maritimes Region, St Andrews Biological Station, 531 Brandy Cove Road, St. Andrews, New Brunswick, Canada E5B 2L9

[1]Bedford Institute of Oceanography,Science Branch, Maritimes Region, Fisheries and Oceans Canada, 1 Challenger Drive, Dartmouth, Nova Scotia, Canada B2Y 4A2

[2]University of Southern Maine, Aquatic Systems Group, 350 Commercial St.Portland, Maine, United States of America 04101

[3]CSIRO Marine & Atmospheric Research, Marine Biodiversity Hub, PO Box 120, Cleveland, Queensland, Australia 4163

This is the two hundred and ninety-thirdTechnical Report
of the Biological Station, St. Andrews, NB

Canadian technical report of fisheries and aquatic sciences

Technical reports contain scientific and technical information that contributes to existing knowledge but which is not normally appropriate for primary literature. Technical reports are directed primarily toward a worldwide audience and have an international distribution. No restriction is placed on subject matter and the series reflects the broad interests and policies of the Department of Fisheries and Oceans, namely, fisheries and aquatic sciences.

Technical reports may be cited as full publications. The correct citation appears above the abstract of each report. Each report is abstracted in *Aquatic Sciences and Fisheries Abstracts* and indexed in the Department's annual index to scientific and technical publications.

Numbers 1-456 in this series were issued as Technical Reports of the Fisheries Research Board of Canada. Numbers 457-714 were issued as Department of the Environment, Fisheries and Marine Service, Research and Development Directorate Technical Reports. Numbers 715–924 were issued as Department of Fisheries and the Environment, Fisheries and Marine Service Technical Reports. The current series name was changed with report number 925.

Technical reports are produced regionally but are numbered nationally. Requests for individual reports will be filled by the issuing establishment listed on the front cover and title page. Out-of-stock reports will be supplied for a fee by commercial agents.

Rapport technique canadien des sciences halieutiques et aquatiques

Les rapports techniques contiennent des renseignements scientifiques et techniques qui constituent une contribution aux connaissances actuelles, mais qui ne sont pas normalement appropriés pour la publication dans un journal scientifique. Les rapports techniques sont destinés essentiellement à un public international et ils sont distribués à cet échelon. Il n'y a aucune restriction quant au sujet; de fait, la série reflète la vaste gamme des intérêts et des politiques du ministère des Pêches et des Océans, c'est-à-dire les sciences halieutiques et aquatiques.

Les rapports techniques peuvent être cités comme des publications complètes. Le titre exact parait au-dessus du résumé de chaque rapport. Les rapports techniques sont résumés dans la revue *Résumés des sciences aquatiques et halieutiques*, et ils sont classés dans l'index annuel des publications scientifiques et techniques du Ministère.

Les numéros 1 à 456 de cette série ont été publiés à titre de rapports techniques de l'Office des recherches sur les pêcheries du Canada. Les numéros 457 à 714 sont parus à titre de rapports techniques de la Direction générale de la recherche et du développement, Service des pêches et de la mer, ministère de l'Environnement. Les numéros 715 à 924 ont été publiés à titre de rapports techniques du Service des pêches et de la mer, ministère des Pêches et de l'Environnement. Le nom actuel de la série a été établi lors de la parution du numéro 925.

Les rapports techniques sont produits à l'échelon régional, mais numéroté à l'échelon national. Les demandes de rapports seront satisfaites par l'établissement auteur dont le nom figure sur la couverture et la page du titre. Les rapports épuisés seront fournis contre rétribution par des agents commerciaux.

Canadian Technical Report of
Fisheries and Aquatic Sciences 2895

2010

A Geodatabaseof Historical and Contemporary Oceanographic Datasets for Investigating the Role of the Physical Environment in Shaping Patterns of Seabed Biodiversity in the Gulf of Maine

by

M.E. Greenlaw, J.A. Sameoto[1], P. Lawton,N.H. Wolff[2], L.S. Incze[2],
C.R. Pitcher[3], S.J. Smith[1] and A. Drozdowski[1]

Fisheries and Oceans Canada, Science Branch, Maritimes Region, St Andrews Biological Station, 531 Brandy Cove Road, St. Andrews, New Brunswick, Canada E5B 2L9

[1]Bedford Institute of Oceanography,Science Branch, Maritimes Region, Fisheries and Oceans Canada, 1 Challenger Drive, Dartmouth, Nova Scotia, Canada B2Y 4A2

[2]University of Southern Maine, Aquatic Systems Group, 350 Commercial St.Portland, Maine, United States of America 04101

[3]CSIRO Marine & Atmospheric Research, Marine Biodiversity Hub, PO Box 120, Cleveland, Queensland, Australia 4163

This is the two hundred and ninety-thirdTechnical Report
of the Biological Station, St. Andrews, NB

© Her Majesty the Queen in Right of Canada, 2010.
Cat. No. Fs 97-6/2895E ISSN 0706-6457 (print version)
Cat. No. Fs 97-6/2895E-PDF ISSN 1488-5379 (online version)

Correct citation for this publication:

Greenlaw, M.E., Sameoto, J.A., Lawton, P., Wolff, N.H., Incze, L.S., Pitcher, C.R., Smith, S.J and Drozdowski, A. 2010.A geodatabase of historical and contemporary oceanographic datasets for investigating the role of the physical environment in shaping patterns of seabed biodiversity in the Gulf of Maine.Can. Tech. Rep. Fish.Aquat. Sci.2895: iv + 35 p.

TABLE OF CONTENTS

Abstract iv
Résumé iv
Introduction 1
CoML Cross-Project Synthesis Description 1
Methods 2
CoML Biodiversity Analysis Cross-Project Methods 2
Source Biological Datasets 2
NOAA Benthic Database Smith-Mcintyre Grab Samples 2
NEFSC Bottom Trawl Survey 5
Selection of Time Periods 7
Geodatabase Creation and Data Access 7
XYZ Text File Creation and Data Access 8
Interpolation Methods 9
Optimal Estimation 9
Spline Interpolation 10
Physical Oceanographic Datasets 10
Bathymetry and Derivatives 10
Substrate 10
Bottom Stress 10
Stratification 11
Benthic Irradiance 11
Chlorophyll 12
Sea Surface Temperature 12
Benthic Temperature 12
Benthic Salinity 13
Dissolved Oxygen 13
Bottom Nutrients 13
Phosphate 13
Silicate 13
Nitrate 14
Discussion 30
Acknowledgements 32
References 32

ABSTRACT

Greenlaw, M.E., Sameoto, J.A., Lawton, P., Wolff, N.H., Incze, L.S., Pitcher, C.R., Smith, S.J and Drozdowski, A. 2010.A geodatabase of historical and contemporary oceanographic datasets for investigating the role of the physical environment in shaping patterns of seabed biodiversity in the Gulf of Maine.Can. Tech. Rep. Fish.Aquat. Sci.2895: iv + 35 p.

The management and conservation of the marine environment and marine resourcesincreasingly requires the synthesis of spatial data from a range of physical and biological features across a variety of scales. The work phase of compiling these data layers for a particular project can be time intensive and costly. However, once these datasets are compiled and processed to generate continuous spatial layers, the preservation of these data in a common georeferenced format can facilitate their use in future work; in particular for spatial planning, decision making and ecosystem-based management. A comprehensive suite of physical, chemical and biological layers (30 layers) for the Gulf of Maine area have been compiled within a single geodatabase, and in xyz format, using publicly-available U.S. and Canadian oceanographic data sources. The primary driver for this effort was the Census of Marine Life's cross-project that involved three regions (Great Barrier Reef, Gulf of Mexico and Gulf of Maine) who investigated the role of physical variables in predicting patterns of biodiversity in seabed assemblages. In this report we provide background methods (including formal metadata), issues faced in compiling geospatial resources, and caveats for subsequent usage for those datasets used in the Gulf of Maine portion of the Census project.

RÉSUMÉ

Greenlaw, M.E., Sameoto, J.A., Lawton, P., Wolff, N.H., Incze, L.S., Pitcher, C.R., Smith, S.J and Drozdowski, A. 2010.A geodatabase of historical and contemporary oceanographic datasets for investigating the role of the physical environment in shaping patterns of seabed biodiversity in the Gulf of Maine.Can. Tech. Rep. Fish.Aquat. Sci.2895: iv + 35 p.

La gestion et la conservation du milieu marin et des ressources marines exigent de plus en plus que l'on procède à une synthèse des données spatiales concernant plusieurs caractéristiques physiques et biologiques à diverses échelles. L'étape de travail consistant à compiler toutes ces couches de données peut être chronophage et coûteuse. Cependant, une fois les données compilées et traitées de manière à générer des couches spatiales continues, la conservation de ces données sous une forme géoréférencée commune peut en faciliter l'utilisation dans des travaux futurs, particulièrement pour ce qui est de la planification spatiale, de la prise de décision et de la gestion écosystémique. On a compilé une série complète de couches de données physiques, chimiques et biologiques (30 couches) sur la région du golfe du Maine en une seule base de données géographiques, en format xyz, en utilisant des sources de données océanographiques des États-Unis et du Canada accessibles au public. Le principal facteur motivant cet effort était le projet conjoint du Census of Marine Life [inventaire de la vie marine] avec trois autres régions (le récif de la Grande Barrière, le golfe du Mexique et le golfe du Maine), projet qui avait pour objectif d'examiner le rôle des variables physiques aux fins de la prévision du profil de la biodiversité des assemblages du fond océanique. Dans le présent rapport, nous présentons la méthodologie contextuelle (notamment les métadonnées officielles), les problèmes rencontrés lors de la compilation de ces ressources géospatiales et les mises en garde relativement à une utilisation ultérieure.

INTRODUCTION

The management and conservation of the marine environment and marine resourcesincreasingly requires the synthesis of spatial data from a range of physical and biological features across a variety of scales. The work phase of compiling these data layers for a particular project can be time intensive and costly. However, once these datasets are compiled and processed to generate continuous spatial layers, the preservation of these data in a common georeferenced format can facilitate their use in future work; in particular for spatial planning, decision making and ecosystem-based management (Fisher and Rahel 2004, Gee 2007, Wood and Dragicevic 2007). In conducting a recent investigation on the role of physical environmental variables in predicting biodiversity composition of benthic and demersal fish and invertebrate assemblages in the Gulf of Maine, we compiled 31 oceanographic layers and seabed characteristics. These layers were then preserved in xyz format and as raster grids in an Environmental Systems Research Institute (ESRI^{TM}) file geodatabase. In this report we provide background methods (including formal metadata), issues faced in compiling these geospatial resources, and caveats for subsequent usage.

CoML Cross-Project Synthesis: Physical Surrogates for Predicting Seabed Biodiversity

In 2000 the International Census of Marine Life (CoML) began a global effort to assess and explain the diversity, distribution and abundance of marine life. This marine research initiative included 14 "field projects" covering different marine regions, habitats and functional groups of organisms within the global ocean (e.g. the Abyssal Plains, Coral Reefs, Zooplankton etc; www.coml.org). One of these field projects, the Gulf of Maine Area program (GOMA) was selected as the CoML's Regional Ecosystem project and focuses on the biodiversity of marine life in the Gulf of Maine (GOM [grey area, Figure 1]). The GOM is located on the easternNorth American continental shelf between 47° and 39°Nlatitude and covers approximately 250K km^2 (see Incze *et al.*(2010) for an overview of the region and its marine biodiversity).

One GOMA initiative has been to participate in a cross-project synthesis with two other regions, the Great Barrier Reef and Gulf of Mexico, to characterize how physical factors affect species distribution and abundance patternsin contrasting ecosystems. The distribution and abundance of marine species and assemblages has been of fundamental interest to science and of considerable importance to management and conservation. For most marine species, such information is severely lacking, partly due to the great expense and time required for ship-based biological surveys. To deal with this problem, methods of generalization are required. Since many benthic organisms are strongly associated with specific habitat characteristics, the CoML project focused on the use of physical environmental variables (e.g. substrate, benthic temperature, nutrient concentrations)to predict the spatial pattern in seabed assemblages.The analysis required the compilation of numerous oceanographic datasets from the GOM. These data were determined to also be important for ecosystem-based studies in general, including those underway with an Ecosystem Research Initiative (ERI; http://www.dfo-mpo.gc.ca/science/publications/fiveyear-plan-quinquennal/index-eng.html#a3_2) being conducted by Fisheries and Oceans Canada (DFO). Therefore, as a prerequisite to the statistical analyses for the seabed diversity project, oceanographic datasets compiled for two major time periods corresponding to the years the biological data were sampled ([1956–1968] and [1996–2007]) were preserved as a collection. These physical datasets, from numerous data sources,were

processed to derive continuous layers for the GOM study area. Biological datasets were assembled from several publicly-available databases as well. However, in this report we only provide information on data processing steps and provide links to the authoritative database providers for each oceanographic dataset. The benthic biological datasets are not included in the compilation as it is anticipated that future studies would proceed by accessing the most up-to-date data directly from these biological databases.

METHODS

CoML Cross-Project Methods

All three regions involved in this project collated broad-scale biological survey datasets comprised of site-by-species abundance data collected from trawls, benthic sleds, and grabs/cores (Figure 1), as well as site-by-physical datasets comprised of available physical variables.

For the GOMA project three large benthic datasets were accessed. These data were originally acquired by federal agencies in Canada (Fisheries and Oceans Canada (DFO)) and the United States (Northeast Fisheries Science Center (NEFSC), National Marine Fisheries Service (NMFS), National Oceanographic and Atmoshpheric Administration (NOAA)).

These datasets represent the three most extensive benthic datasets in the GOM region: the NEFSC Bottom Trawl Surveys, DFO Ecosystem Surveys, and the NEFSC Benthic Database (Smith-McIntyre Grab samples). Each dataset was filtered to only include taxa identified to genus or species.

Oceanographic habitat variables for the time period of each biological dataset were collected and included: bathymetry and derivatives, seabed current stress, sediment characteristics, benthic irradiance, nutrients, temperature, salinity and chlorophyll. The oceanographic habitat variables along with the biological data were analyzed using Gradient Forest (a modified version of a Random Forests analysis (Breiman 2001)) to identify important environmental variables influencing the distribution and abundance of benthic species in the GOM. A more complete summary of the Gradient Forest statistical methods is provided by Ellis *et al.* (2010 In Prep).

Source Biological Datasets

NOAA Benthic Database Smith-McIntyre Grab Samples

The NOAA Benthic Database was accessed prior to its release to the Ocean Biogeographic Information System (OBIS, http://www.iobis.org/), although, by this time virtually all quality control measures had been completed. Only Smith-McIntyre Grab samples were selected that occurred between 1956–1968, inclusive, in the GOMArea (Theroux and Wigley 1998). This time period was selected to include the majority of the data collected from Smith-McIntyre Grab samples, collected between 1956–1985 (Figure 2). The Smith-McIntyre spring-loaded grab sampled 0.1 m^2 of bottom area and had a volume of approximately 15 L. Sampling primarily took place during summer months with the majority of collections made in July and August. Samples (n=618) were distributed across all years,1956–1968, although 85% were collected from 1957–1960 (Figure 2). Analyses included 315 species from 478 stations that coincided with all the physical variables, including between 1 and 25 replicates per location.The main taxa in this dataset included Arthropoda (177) and Mollusca (162), although annelids,

echinoderms, bryozoans, sipunculids, cnidarians, hemichordates and baccilariophytanswere also present with the number of taxa ranging from 1 to 28. The geographic distribution of samples in the Gulf of Maine excluded the western Scotian Shelf and the Bay of Fundy (Figure 1).

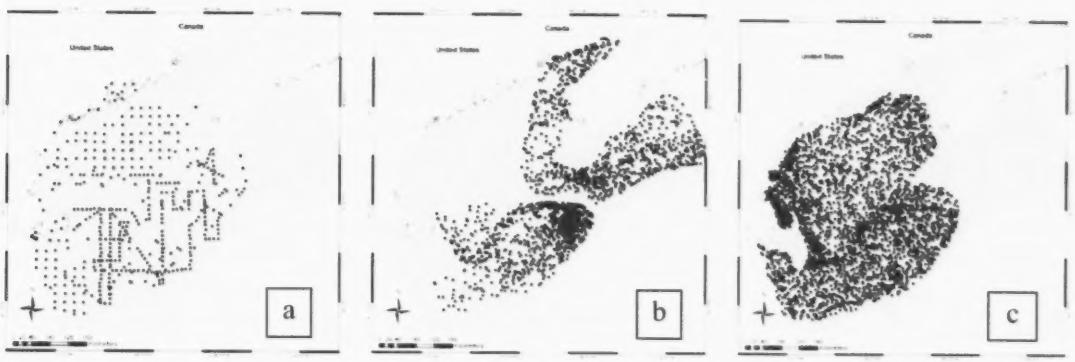


Figure 1. Distribution of biological samples in the Gulf of Maine (grey shaded area denotes the geographic region of interest for CoML'sGOMA program) (a) Smith-McIntyre Grab Samples collected by Theroux and Wigley(1998) between 1956-1968 (b) DFO Winter and Summer Ecosystem Surveys stations from 2000-2007 (c) NMFS Benthic Trawl Surveys from 1997-2007.

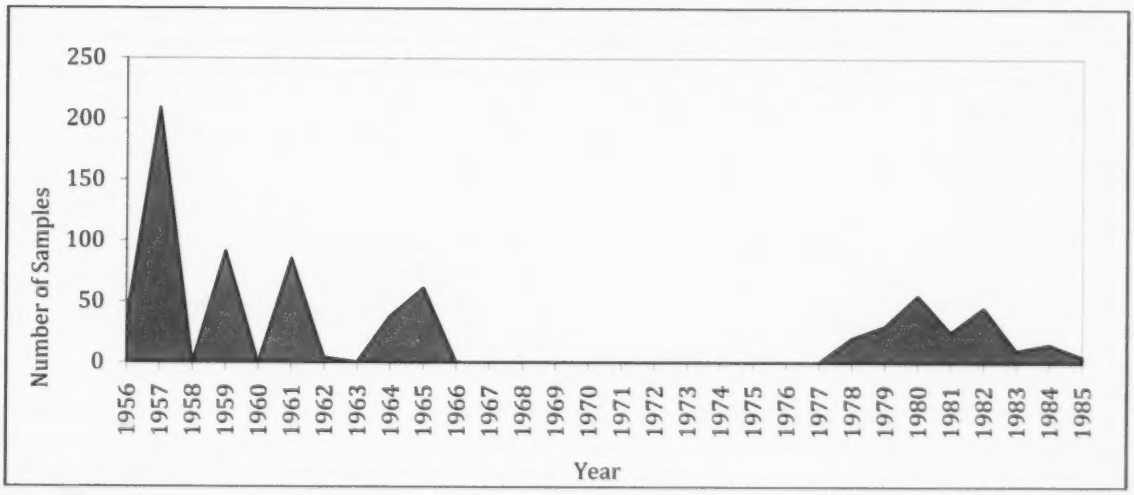


Figure 2. Number of Smith-McIntyre grab samples collected per year (Theroux and Wigley 1998).

Data collected during the DFO Ecosystem Surveys were exported from the DFO Groundfish database for trawls conducted on the Scotia Shelf, in the Bay of Fundy or on Georges Bank from 1996–2007, inclusive. The DFO Annual Ecosystem Survey samples demersal fish and has increasingly been recording data on invertebrate species. The Scotian Shelf surveys were first implemented in 1970 and occur in the summer (July) while the Georges Bank surveys began in 1986 and take place in Feb–March (Clark 2010).

The survey uses a stratified random sampling design where stratification is primarily based on depth, and the allocation of the number of stations per stratum is proportional to the variance in catch of haddock (Clark, 2010). The Scotian Shelf summer survey contains 48 strata (Figure 3) distributed among four depth zones (< 92 m, 93 – 183 m, > 184 m, and mixed) to a maximum of 731 m along the shelf edge (Frank 2004). The Georges Bank spring survey is stratified among 3 depth zones (< 93 m, 93 – 182, and mixed). The same trawl gear, the Western IIA, was used for surveys during the 1996 to 2007 period. However, irregularities in the rigging of the trawl in the 2004 summer survey made direct comparisons of catch for some species questionable and this survey was excluded from the analysis (Clark 2010). The trawl gear uses a small mesh codend liner (19mm) capable of retaining forage and small, non-commercial species. At each sampling station within each stratum, a standard bottom tow defined to be a 30 minute haul on bottom at 3.5 knots was conducted. This results in an area swept of 0.0404 km^2 for a standard tow(Shackell and Frank 2003).More details on the surveys can be found in Chadwick *et al.* (2007).

The original data were exported from a dataset published on the Ocean Biogeographic Information System, OBIS (Clark and Branton 2007) for the period from 1996–2007, and included 1032 trawls on the Scotian Shelf and 1299 trawls on Georges Bank (Figure 2). The spring survey included 112 fish species comprised of 99 genera, and 49 invertebrate species comprised of 49 genera.

The most frequent species in the Scotian Shelf summer survey dataset included: *Melanogrammusaeglefinus* (Haddock), *Clupeaharengus* (Herring), *Merlucciusbilinearis* (Silver Hake), *Squalusacanthias* (Spiny Dogfish), *Illexillecebrosus* (Short-finned Squid), *Myoxocephalusoctodecemspinosus* (Longhorn Sculpin), *Pseudopleuronectesamericanus* (Winter Flounder), *Pollachiusvirens* (Pollack), *Placopectenmagellanicus* (Sea Scallop), and *Limandaferruginea* (Yellowtail Flounder).

In the Georges Bank spring survey dataset the most frequent species included: *Melanogrammusaeglefinus* (Haddock), *Clupeaharengus* (Herring), *Myoxocephalusoctodecemspinosus* (Longhorn Sculpin), *Limandaferruginea* (Yellowtail Flounder), *Merlucciusbilinearis* (Silver Hake), *Scomberscombrus* (Atlantic Mackerel), *Illexillecebrosus* (Short-finned Squid), *Leucorajaocellata*(Winter Skate),*Gadusmorhua*(Cod), and *Leucorajaocellata* (Winter Skate)

With both datasets analyses were performed with and without data collected before 2000. These data were excluded as some shrimps, crabs and scallops were only recorded since 1999 (Tremblay *et al.* 2007, Clark 2010). For the Scotian Shelf summer survey dataset invertebrates includedin the analysis were: *Cancer borealis, Cancer irroratus, Chaceonquinquedens, Chionocetesopilio, Homarusamericanus, Hyasaraneus, Hyascoarctatus, Lithodesmaja, Pandalus borealis, Pandalusmontagui, Illexillecebrosus*and*Placopectenmagellanicus*. Abundances of *Pandalus borealis* and *Pandalusmontagui* were determined by dividing by the average species weight as only the weight was reliably measured.For the Georges Bank spring survey dataset the invertebrate taxa included in the analysis were only *Homarusamericanus* and *Illexillecebrosus*.

The final analysis only included taxa that were recorded in > 5% of samples.Therefore, rare species were not included in the analysis—a more detailed description of the reasoning and implications of this decision be found in Ellis *et al.* (2010 In Prep). Samples were also removed that did not have corresponding physical data in any one of the physical variables. The RandomForest analyses included 81 species in the late winter-early spring survey, and 95 species in the summer survey.

NEFSC Bottom Trawl Survey

NEFSC of NMFS conducts trawl surveys in the GOM in both the fall and spring using a stratified random sampling design. The fall surveys began in 1963 and sample depths from 27 to 365 m (Despres-Patanjo *et al.* 1988). The spring survey series began in 1968. In 1972 the geographic coverage of the surveys were extended to inshore areas landward of the 27 m isobath. The stratified random sampling assures a fairly uniform distribution of stations throughout the survey areas with an average allocation per seasonal survey of 350 stations. Strata are delineated by depth (Figure 3). Stations were allocated to strata in proportion to strata area and were randomly assigned to specific locations within strata (Azarovitz 1981). The Yankee 36 Bottom Trawl used for the survey sweeps 0.0334 km^2 during a standard 30 min tow at 3.8 knots, which is slightly less than the area swept in the DFO Ecosystem survey.

Data from the fall and spring surveys were exported (and treated separately) for the 1996–2007 time period. This included 2001 tows that were done in the fall (average of 164 per year, min. 152, max. 192) and 1975 tows that were done in spring (average of 165 per year, min156, max 182). The fall surveys generally take place in October (1643 tows) but occasionally occurs in September (226 tows) and December (132 tows), while the spring survey usually takes place in April (1723 tows) with occasional tows in March (238 tows) and May (14 tows). The data used are available as an OBIS dataset (NOAA's National Marine Fisheries Service Northeast Fisheries Science Center 2005), that was clipped to the GOM area and filtered to only include taxa identified to genus or species. Since the inception of these surveys in 1963, species identifications have increased to include more invertebrate data, and our selection of survey data from 1996 onwards was in part due to the species identifications for fish (fall: 146 species, 124 genera; spring: 98 species, 88 genera) having reaching 100%.Invertebrates in the dataset included 33 species and 29 genera in the fall and 26 species and 23 genera in the spring. Data on invertebrate species has increasingly been collected and recorded for this survey. Data from the fall and spring surveys were treated separately for the 1996–2007 time period. Samples were also removed that did not overlap with corresponding physical data in any one of the physical variables.

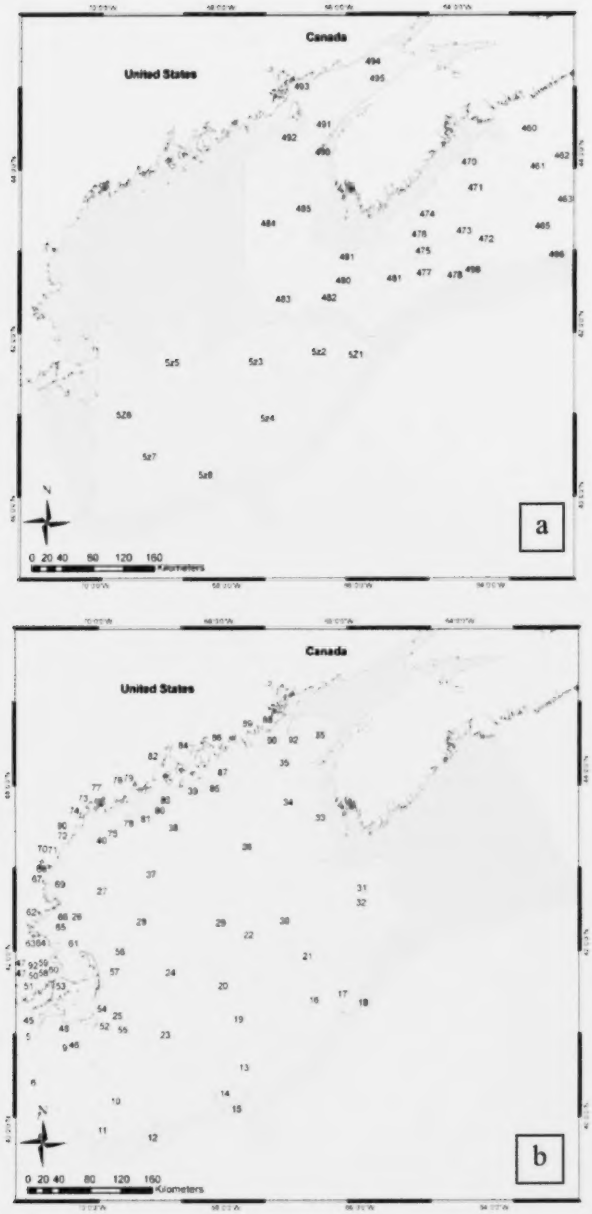


Figure 3.(a) The DFO Ecosystem Survey strata for Georges Bank and the Scotian Shelf in the GOM area (grey area). Labels indicate strata names (b) The NMFS Bottom Trawl Survey strata in the GOM area. Labels indicate strata names.

Selection of Time Periods

Time periods of the biological and corresponding physical data were selected based on those years that had abundant data available for each of the biological datasets selected. The Smith-McIntyre grab samples included data from 1956–1985, however most data was collected from 1956–1968 and therefore only those years were included in the analyses. Corresponding physical data for the 1956-1968 timeperiod was largely present, as most time series datasets for oceanographic variables began in the 1930s; excluding silicate which only began being recorded in 1961. Satellite information was only available for the contemporary time period as the AVHRR and SeaWifs satellites only began collecting data in the early to mid-1990s. However, it was decided to use the contemporary satellite for the 1956–1968 time period as no other data were available. Itwas also determined that the 1956–1968 time period did not have enough nitrate and silicate data to warrant creating layers.

The time periods selected for the DFO Ecosystem Surveys and the NEFSC Bottom Trawl Surveys datasets were to be from 1996–2007; however, it was decided to run analyses not including years where invertebrates had not been rigorously recorded in the DFO Ecosystem Surveys (before the year 2000). Oceanographic datasets were largely available for this time–period, although satellite sea surface temperate and chlorophyll only began to be sampled in 1997, and dissolved oxygen was not available for this period from the GOM Region Nutrient and Hydrographic database.

Geodatabase Creation and Data Access

The physical environmental variables were preserved in an ESRI (Environmental Systems Research Institute, RedlandsCalifornia) file geodatabase that can be accessed by contacting Michelle Greenlaw (St. Andrews Biological Station, michelle.greenlaw@dfo-mpo.gc.ca). A geodatabase is a database designed to store, query and manipulate geographic information and spatial data. All data are stored in raster format (coordinate system: World Geodetic System of 1984, Universal Transverse Mercator Zone 19 [WGS84 UTM Zone 19]). Each layer also has an associated metadata file which includes the data source, description, purpose, supplementary information and contact information for the layer. Metadata can be accessed by opening the layer in ESRI's ArcCatalog and using the Metadata tab (Figure 4).

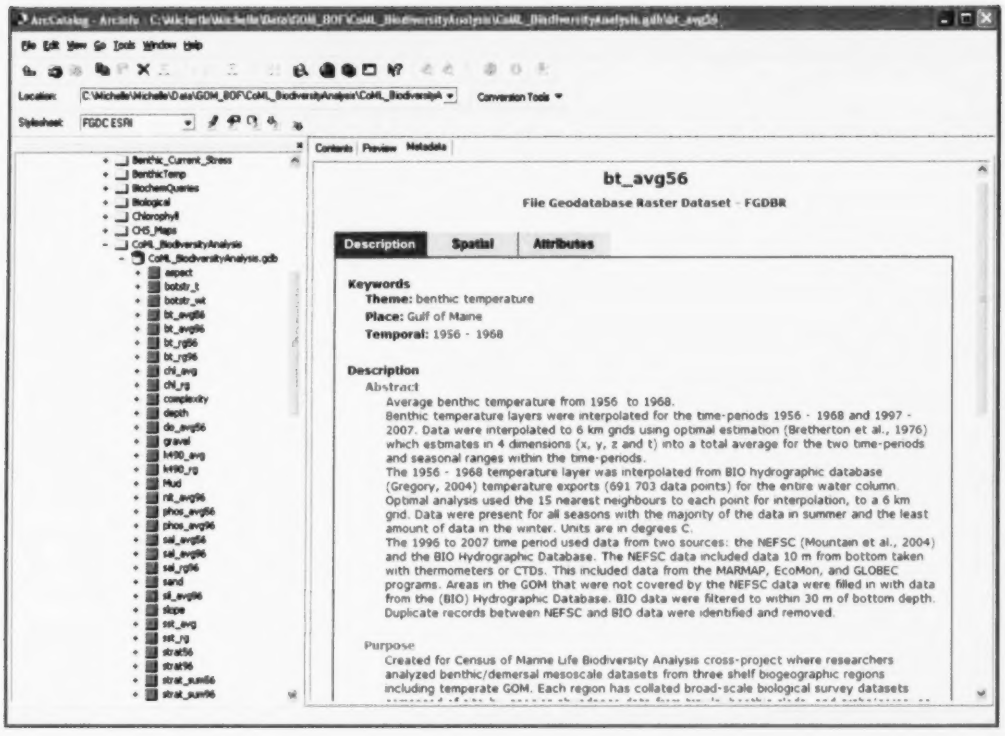

bt_avg56
File Geodatabase Raster Dataset - FGDBR
Description
Spatial
Attributes
Keywords
Theme: benthic temperature
Place: Gulf of Maine
Temporal: 1956 - 1968
Description
Abstract
Average benthic temperature from 1956 to 1968.
Benthic temperature layers were interpolated for the time-periods 1956 - 1968 and 1997 - 2007. Data were interpolated to 6 km grids using optimal estimation (Bretherton et al., 1976) which estimates in 4 dimensions (x, y, z and t) into a total average for the two time-periods and seasonal ranges within the time-periods.
The 1956 - 1968 temperature layer was interpolated from BIO hydrographic database (Gregory, 2004) temperature exports (691 703 data points) for the entire water column. Optimal analysis used the 15 nearest neighbours to each point for interpolation, to a 6 km grid. Data were present for all seasons with the majority of the data in summer and the least amount of data in the winter. Units are in degrees C.
The 1996 to 2007 time period used data from two sources: the NEFSC (Mountain et al., 2004) and the BIO Hydrographic Database. The NEFSC data included data 10 m from bottom taken with thermometers or CTDs. This included data from the MARMAP, EcoMon, and GLOBEC programs. Areas in the GOM that were not covered by the NEFSC data were filled in with data from the (BIO) Hydrographic Database. BIO data were filtered to within 30 m of bottom depth. Duplicate records between NEFSC and BIO data were identified and removed.
Purpose
Created for Census of Marine Life Biodiversity Analysis cross-project where researchers analyzed benthic/demersal mesoscale datasets from three shelf biogeographic regions including temperate GOM. Each region has collated broad-scale biological survey datasets

Figure 4. A view of ESRI®'s ArcCatalog showing the geodatabase in the left panel, expanded to show each physical layer. The right panel is a view of the metadata for the benthic temperature layer (1956–1968).

XYZ Text File Creation and Data Access

Physical environmental variables were also exported in xyz format (three column format including latitude, longitude then physical variable) that can be accessed by contacting Michelle Greenlaw (St. Andrews Biological Station, michelle.greenlaw@dfo-mpo.gc.ca). This format is easily imported by all Geographic Information Systems (GIS) software and is more accessible to certain non-GIS software programs. The files are in the World Geodetic System of 1984 (WGS84) coordinate system. The xyz files do not have associated metadata therefore this report or the geodatabase will serve as metadata for those files.

Interpolation Methods

Optimal Estimation

The program OAX version 5.1 was used to interpolate continuous data layers for a number of point-based oceanographic datasets. This program, developed in the early 1990's at DFO(He *et al.* 2003),applies the method of optimal interpolation to estimate the values of variables at specified points in space and time. Each interpolated point is calculated using a nearest neighbour algorithm, where the weighted average is taken for a specified number of data points that are closest to the interpolated point. Unlike many other interpolation techniques, OAX allows the user to interpolate in four dimensions: longitude (x), latitude (y), depth (z), and time (t).

To run OAX, the program and three additional files are required; a grid, deck, and data file. The grid file is a text file containing a list of grid points. Each point is where an estimated data value will be calculated. Optimally estimated benthic data layers were created using a grid file derived from the United States Geological Survey (USGS) North American 15 arc-second Digital Elevation Model clipped to the GOM extent(Roworth and Signell 1998). This grid was resampled to coarser resolutions when required due to the spatial distribution of sample coverage of the oceanographic datasets.

The deck file is a text file that specifies the dependent variable to be estimated, the independent variables (i.e. x, y, z, t), the data file to be interpolated, the grid file, the number of nearest neighbour data points to use, the global scales, and the statistical model to be used. The global scales are the overall scales for the independent variables that are used to weight the independent variables and define a distance calculation. This distance enables the selection of the closest number of nearest neighbours from which a value will be interpolated for each grid point. The number of nearest neighbours was a function of the spread of the data points for each interpolated layer. For all runs of OAX, the statistical model was the estimated mean.

The data file is a text file that contains the dataset to be interpolated. All data in both the grid and data files must be in a Cartesian coordinate system. Data was reprojected from the geographic coordinate systems NAD 83 or WGS 84 to the projected coordinate system WGS 84 UTM Zone 19. This resulted in equivalent units (meters) in all spatial directions (x, y, and z); which is a requirement for OAX to perform its mathematical calculations.

To derive seasonal data layers, input data were restricted to correspond to the following day-of-year limits:

Season	Day-of-Year Limits
Winter	1–90
Spring	91–181
Summer	182–273
Fall	274–365

The code for OAX 5.1 was compiled for windows and can be run through the MS-DOS command window. Additional information and documentation on OAX can be found online at http://www2.mar.dfo-mpo.gc.ca/science/ocean/coastal_hydrodynamics/Oax/oax.html.

Spline Interpolation

The ESRI ArcGIS Spatial Analyst extension was used for spline interpolations of substrate and stratification samples as these layers did not require interpolation in 4 dimensions (x, y, z and t). Parameters for each specific interpolation are listed in the description of the physical environmental dataset. Spline interpolation is a method that estimates values using a mathematical function that minimizes overall surface curvature, resulting in a smooth surface that passes exactly through the input point. Spline was chosen as the interpolation method to ensure the fitted surface passes through the input points, as many of the substrate and stratification points provided were taken at the same time as the biological samples. This method would ensure the accuracy of the estimate at those points. At the points where an actual sample of substrate or stratification was not available, the spline method has an average accuracy in comparison with other interpolation methods. The average relative mean absolute error and relative root mean square error were 40% and 30% when tested, respectively (Li and Heap 2008).

Physical Oceanographic Datasets

Bathymetry and Derivatives

AUSGS digital elevation model (Roworth and Signell 1998) at a resolution of 15 arc-seconds or 650 m (Figure 5), was used to intersect biological samples with bathymetry information and to calculate slope (maximum change in depth in the 8 surrounding grid cells, degrees), aspect (degrees from north),bathymetric position index (BPI, unitless), and benthic complexity (maximum change in slope in the 8 surrounding grid cells, degrees). The data unit of the bathymetry layer was in meters.

Substrate

Sediment values were extracted from the standing stock ofUSGS (Poppe *et al.* 2005) and Canadian Geological Survey (Geological Survey of Canada 2009)database records (US: 1955–2004, Can: 1964–2003). These were combined into one point layer of sediment samples. Fields in the data includedpercent sand, percent mud, and percent gravel from point samples (Figure 6). Point datawere interpolated to a 6000 m resolution raster grid using the spline method. The spline parameters used were: Tense, 40 Weight, and a 3 point average. Resolution of the output layers was determined by the density of the original data.

Bottom Stress

Bottom stress (Figure 7) was calculated from the frictional velocity maps of Drozdowski and Hannah (2010), which captured the effect of waves and modeled tidal bottom currents. Their friction velocity calculation was done on a grid covering the GOM and Scotian Shelf and was the synthesis of 3 earlier data products: (1) the high resolution bathymetry data (0.25 min for the GOM and 0.4 min for the Scotian Shelf), (2) 42 year hindcast of the wave height and period data (Swail and Cox 2000), and (3) the near-bottom tidal currents obtained from a combination of 3D model and 2D tidal model results of Hannah *et al.* (2001) and Han and Loder(2003). The final calculation of friction velocity was performed using the sediment transport model SEDTRANS96 (Li and Amos 2001). Results of the 90th percentile significant wave height and period were used, which have the interpretation of representing moderate to large wave events that occur 10% of the time (~one month per year).

Frictional velocity was converted to bottom stress using the formula (Condie and Webster 1997): bottom stress= (Bottom frictional velocity) ^{2} * water density where, water density = 1027.5 kg/m^3. The output resolution of the raster data layer was 952 m.

Bottom stress with only the influence of tides (Figure 7) was calculated using frictional velocity in m^2s^{-2} exported from Gulf of Maine Ocean Observation System (GOMOOS) Nowcast Forecast System (Xue*et al.* 2005). Frictional velocity was multiplied by water density (1027.5) to calculate bottom stress. Bottom stress in the GOM is driven mostly by tidal flow therefore; the largest source of temporal variability is from the lunar cycle. A layer of bottom stress was calculated for each complete month (to capture lunar cycle) rather than a year or more since monthly data captures the majority of temporal and spatial variability in the GOM. The frictional velocity vector data received was calculated every 3 hours from August 1, 2008 to August 31, 2008 (248 time layers). After converting the vector data to scalar data, summary statistics were calculated for each model grid point as follows (n = 248): Mean; STDDev; Min.; Max.; Range (Max-Min). These summary statistics were conserved in a GIS polygon layer. The data were then converted to a raster at a resolution of 3800 m.

Stratification

Stratification layers (Figures 8 and 9) were calculated using density point samples from the DFO Hydrographic (Climate) database (Gregory 2004) as the densitydifference between the surface (0 m) and 50 m (Helbig and Higdon 2009). Density values were exported at 0 and 50 m depth for the time periods 1956–1968 and 1996–2007 and from both January to December (yearly average layer) and May to September (summer average layer). Points were used only if they had a corresponding value at both 0 and 50 m depth. The resulting points were interpolated to 2500 m grids using the spline method with the parameters: Tense, 40 Weight, 3 point average.

Sampling effort and its monthly distribution varied between the two time periods. The 1956–1968(Figure 8) time-period included 8654 more density samples than the 1996–2007 time-period (Figure 9). From1956–1968,most samples were collected in August (14% of samples), while September had the least sampling effort (4.5% of samples). During this time period, 1968 was the most sampling intensive year, comprising 15% of all samples, while 1961 was the least sampling intensive year with 2% of all samples. For the 1996–2007 time period, April had the most samples (15% of samples), while December had the fewest (0.9% of samples). Most samples were collected in 1998 (13% of samples), while 2003 had the least number of samples (5% of samples). The resolution of the output layers was determined by the original data density.

Benthic Irradiance

SeaWiFS K490 data (Figure 10) were downloaded for the time–period of 1997–2008, from the NOAA OceanColor ftp download site (NASA 2007) at an 8000 m resolution. K490 indicates the turbidity of the water column; how visible light in the blue–green region of the spectrum penetrates within the water column. Monthly composites were used to calculate a value corresponding to the month each benthic sample was taken. These values were included in the formula below (as K490_monthN) used to calculate a seasonal benthic irradiance:

Benthic Irradiance = Cos ((LAT-offsetN) / 180 * π) * Exp (K490_monthN * Depth)

Where offsetN is the position of the sun for month N:

MonthOffsetN	
January	-21.2
February	-12.2
March	0.0
April	12.3
May	21.2
June	24.5
July	21.2
August	12.2
September	0.0
October	-12.3
November	-21.2
December	-24.5

Chlorophyll

Chlorophyll (CHL, mg m^{-3}) data (Figure 10) were acquired from the Satellite Oceanography Laboratory, University of Maine (PI: Andrew Thomas)who downloads and processes Sea-viewing Wide Field-of-View Sensor (SeaWiFS) CHL data. These data included monthly composites for the time-period of 1997–2008 and were used to calculate a yearly average and range at a 855 m resolution.

Sea Surface Temperature

Sea surface temperature (SST, degrees C) data (Figure 10) were acquired from the Satellite Oceanography Laboratory, University of Maine (PI: Andrew Thomas), who downloads and processes SST data from the Advanced Very High Resolution Radiometer (AVHRR) satellite. These data include monthly composites for the 1997–2008 time-period. These data were then used to calculate the yearly average and seasonal range over the time-period. The data were available at a resolution of 972 m.

Benthic Temperature

Benthic temperature layers were derived for the time-periods 1956–1968 (Figure 11), and 1996–2007 (Figure 12). The 1956–1968 time period used data exported from the DFOHydrographic (Climate) database for the entire water column (Gregory 2004) (n = 691 703 data records). The 1996–2007 time period used data from two sources: the NEFSC (Mountain *et al.* 2004) and the DFO Hydrographic (Climate) database. The NEFSC data included data 10 m from bottom taken with thermometers or Conductivity–Temperature–Depth recorders (CTD). This included data from the NEFSC Bottom Trawl Surveys, Marine Resources Monitoring Assessment and Prediction (MARMAP),Ecosystem Monitoring (EcoMon), and GlobalOcean and Ecosystem Dynamics (GLOBEC) programs. Areas in the GOM that were not covered by the NEFSC data were filled in with data from the DFO Hydrographic (Climate) Database. Hydrographic database data were filtered to within 30 m of bottom depth. Duplicate records between NEFSC and BIO data were identified and removed. The final dataset consisted of 18132 data records. For both time periods, data were present for all years and all seasons with the majority of the data in summer and the least amount of data in the winter.

Optimal estimation routines were used to derive seasonal average temperature layers for each time period at a 6 km resolution, using 15, and 10 nearest neighboursfor the 1956–1968 period, and 1996–2007 period, respectively. These data were then converted into a raster grid. From the seasonal averages a total average and seasonal range was calculated. The data units of these layers are degrees Celsius. The resolution of the output layers was determined by the original data density.

Benthic Salinity

Salinity layers (psu) were created for two time periods: 1956–1968 (Figure 13) and 1996–2007 (Figure 14). For the 1956–1968 salinity layer, data were extracted from the DFO BioChem database of biological and chemical oceanographic data (Gregory and Narayanan 2003, Fisheries and Oceans Canada 2006). Data were filtered to within 20 m of bottom and optimal estimation routines were used to derive an annual average layer at a resolution of 40 km using the 6 nearest neighbours. A total of 408 points were used. There was not enough data within each season to calculate seasonal averages or an overall seasonal range.

For the 1996–2007 time period, two sources of data were used, the NEFSC dataset and the DFO Hydrographic (Climate) Database. Where there were spatial gaps in the NEFSC data, data from the DFO Hydrographic (Climate) Database were used. Data from NEFSC were filtered to within 10 m of bottom while data from the DFO Hydrographic (Climate) Database were filtered to within 30 m of bottom. Duplicates between the two datasets were removed and the total number of salinity records was 14297. Each season had a minimum of 2000 data records. Optimal estimation was used to derive seasonal benthic salinity layers using the 10 nearest neighbors at a resolution of 6 km. The seasonal layers were then used to create total average and seasonal range benthic salinity data layers. The resolution of the output layers was determined by the original data density.

Dissolved Oxygen

Dissolved oxygen (Figure 15) was exported from the DFO BioChem Database for the time period 1956–1968. Values were filtered to within 20m off bottom for a total of 330 dissolved oxygen records. Optimal estimation was used to derivean annual average benthic dissolved oxygen layer using the 6 nearest neighbours at a resolution of 40 km. There was not enough data within each season to calculate seasonal averages or an overall seasonal range. Dissolved oxygen data were not available from the GOM Region Nutrient and Hydrographic database (Rebuck et al. 2009) for the 1996–2007 time period. Data are in micromoles per litre.

Bottom Nutrients

Phosphate: Phosphate (μm) was exported from the DFO BioChem Database for the time period of1956–1968. Values were filtered to within 20 m off bottom for a total of 196 phosphate records. Optimal estimation was used to derivean annual average benthic phosphate layerusing the 6 nearest neighbours at a resolution of 40 km (Figure 16). There was not enough data within each season to calculate seasonal averages or an overall seasonal range.

Phosphate data from 1996–2007 were exported from the GRAMPUS and Hydrographic Databases (Rebuck*et al.*, 2009) for a total of 52 833 observations, from the entire water column. Optimal estimation was used to derivean annual average benthic phosphate layer using the 15 nearest neighboursat a resolution of 6 km (Figure 17). The spatial coverage of the data was not sufficient for seasonal layers to be derived.

Silicate: Silicate data (μm) from 1996–2007 were exported from GRAMPUS for a total of 53 500 observations, from the entire water column. Optimal estimation was used to derivean

annual average silicate layer using the 15 nearest neighboursat a resolution of 6 km (Figure 18). The spatial coverage of the data was not sufficient for seasonal layers to be derived.

Nitrate: Nitrate data (μm) from 1996–2007 were also exported from GRAMPUS for a total of 53 500 observations, for the entire water column. Optimal estimation was used to derivean annual average nitrate layer using the 15 nearest neighboursat a resolution of 6 km (Figure 19). The spatial coverage of the data was not sufficient for seasonal layers to be derived.

Table 1.Physical environmental variables with a GOM-scale coverage used for analysis of the three regional biological datasets referred to in text. (SST = Sea Surface Temperature, K490 =Mean Diffuse Attenuation Coefficient, Stratification = density difference between 0m and 50m depth, HD = DFO Hydrographic (Climate) Database, USGS = United States Geological Service, SeaWifs = Sea-viewing Wide Field-of-view Sensor, CGS = Canadian Geological Survey, NEFSC = Northeast Fisheries Science Centre).

Variable	Geodatabase Name	Units	Resolution (m)	Data Source
Aspect	ASPECT	degree	650	USGS
Benthic Current Stress with Wind and Tidal Influences	BOTSTR_WT	newtons meters $^{-2}$	971	Model
Benthic Current Stress with only tidal influence	BOTSTR_T	newtons meters $^{-2}$	3800	Model
Average Benthic Temperature 1956–1968	BT_AVG56	degree C	6000	HD
Average Benthic Temperature 1996–2007	BT_AVG96	degree C	6000	NEFSC/HD
Seasonal Range of Benthic Temperature 1956–1968	RG_BT56	degree C	6000	HD
Seasonal Range of Benthic Temperature 1996–2007	RG_BT96	degree C	6000	NEFSC/HD
AverageSea Surface Chlorophyll	CHL_AVG	mg m^{-3}	855	SeaWifs
Seasonal Range of Sea Surface Chlorophyll	CHL_RG	mg m^{-3}	855	SeaWifs
Benthic Complexity	COMPLEX	degree	650	USGS
Depth	DEPTH	meters	652	USGS
Benthic Dissolved Oxygen 1956–1968	DO_AVG56	μM	40000	Biochem
Gravel	GRAVEL	percent	6000	USGS/CGS
Average K490	K490_AVG	none	8000	SeaWifs
Seasonal Range of K490	K490_RG	none	8000	SeaWifs
Mud	MUD	percent	6000	USGS/CGS
Benthic Nitrate 1996–2007	NIT_AVG96	μM	40000	GRAMPUS
Benthic Phosphate 1956–1968	PHOS_AVG56	μM	40000	Biochem
Benthic Phosphate 1996–2007	PHOS_AVG96	μM	40000	GRAMPUS
Benthic Salinity 1956–1968	SAL_AVG56	psu	6000	Biochem
Benthic Salinity 1996– 2007	SAL_AVG96	psu	6000	HD/NEFSC
Seasonal Range of Benthic Salinity 1996–2007	SAL_RG96	psu	6000	HD/NEFSC
Sand	SAND	percent	6000	USGS/CGS
Benthic Silicate	SIL_AVG96	μM	6000	GRAMPUS
Slope	SLOPE	degree	650	USGS
Average SST	SST_AVG	degree C	972	SeaWifs
Seasonal Range of SST	SST_RG	degree C	972	SeaWifs
Stratification from 1956–1968	STRAT56	none	2500	HD
Stratification from 1996–2007	STRAT96	none	2500	HD
Summer Stratification from 1956–1968	STRAT_SUM56	none	2500	HD
Summer Stratification from 1996–2007	STRAT_SUM96	None	2500	HD

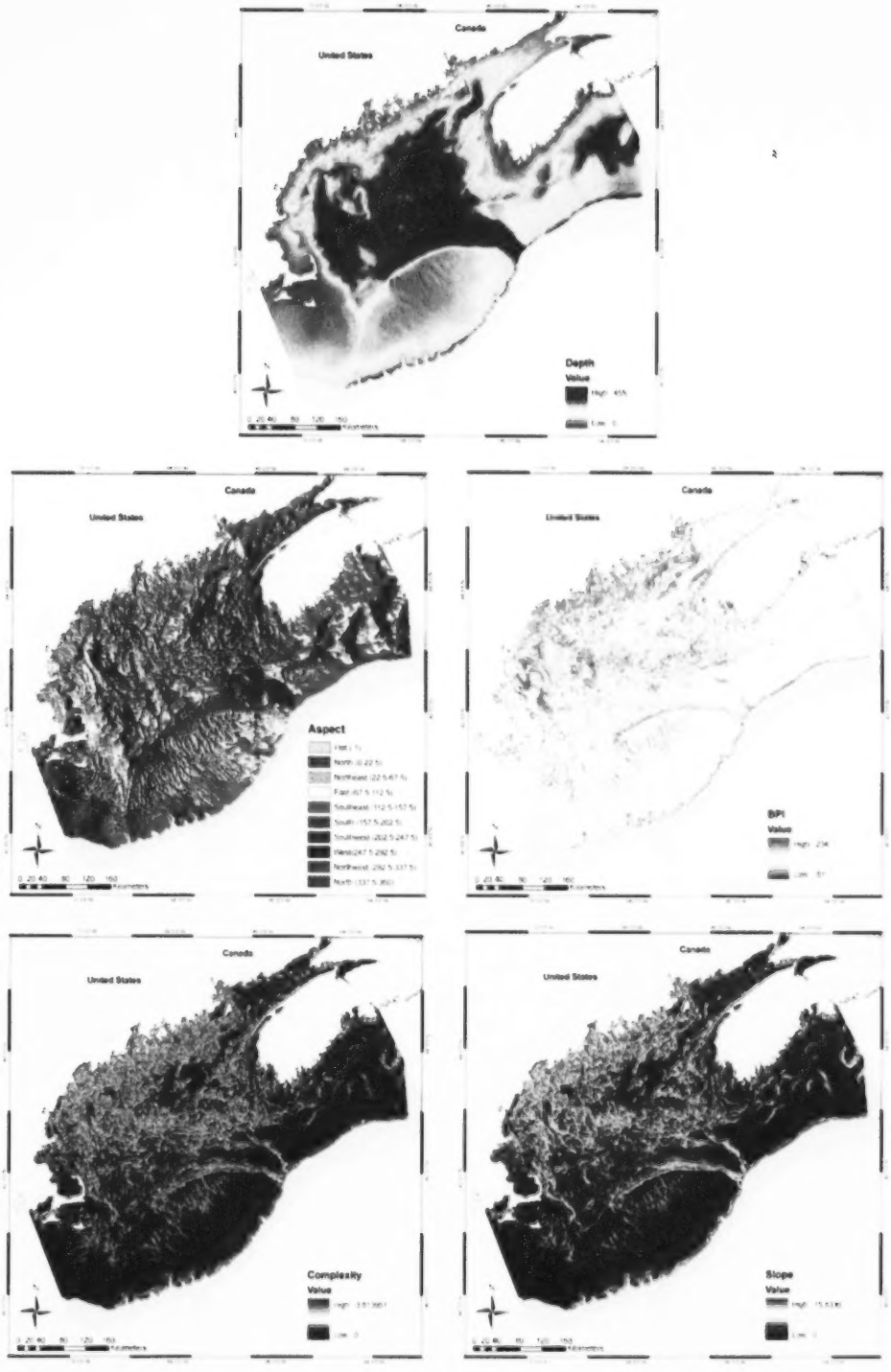


Figure 5. Depth and derivatives created from the USGS 15 arc-second digital elevation model of the GOM.

annual average silicate layer using the 15 nearest neighboursat a resolution of 6 km (Figure 18). The spatial coverage of the data was not sufficient for seasonal layers to be derived.

Nitrate: Nitrate data (μm) from 1996–2007 were also exported from GRAMPUS for a total of 53 500 observations, for the entire water column. Optimal estimation was used to derivean annual average nitrate layer using the 15 nearest neighboursat a resolution of 6 km (Figure 19). The spatial coverage of the data was not sufficient for seasonal layers to be derived.

Table 1.Physical environmental variables with a GOM-scale coverage used for analysis of the three regional biological datasets referred to in text. (SST = Sea Surface Temperature, K490 =Mean Diffuse Attenuation Coefficient, Stratification = density difference between 0m and 50m depth, HD = DFO Hydrographic (Climate) Database, USGS = United States Geological Service, SeaWifs = Sea-viewing Wide Field-of-view Sensor, CGS = Canadian Geological Survey, NEFSC = Northeast Fisheries Science Centre).

Variable	Geodatabase Name	Units	Resolution (m)	Data Source
Aspect	ASPECT	degree	650	USGS
Benthic Current Stress with Wind and Tidal Influences	BOTSTR_WT	newtons meters $^{-2}$	971	Model
Benthic Current Stress with only tidal influence	BOTSTR_T	newtons meters $^{-2}$	3800	Model
Average Benthic Temperature 1956–1968	BT_AVG56	degree C	6000	HD
Average Benthic Temperature 1996–2007	BT_AVG96	degree C	6000	NEFSC HD
Seasonal Range of Benthic Temperature 1956–1968	RG_BT56	degree C	6000	HD
Seasonal Range of Benthic Temperature 1996–2007	RG_BT96	degree C	6000	NEFSC HD
AverageSea Surface Chlorophyll	CHL_AVG	mg m^{-3}	855	SeaWifs
Seasonal Range of Sea Surface Chlorophyll	CHL_RG	mg m^{-3}	855	SeaWifs
Benthic Complexity	COMPLEX	degree	650	USGS
Depth	DEPTH	meters	652	USGS
Benthic Dissolved Oxygen 1956–1968	DO_AVG56	μM	40000	Biochem
Gravel	GRAVEL	percent	6000	USGS CGS
Average K490	K490_AVG	none	8000	SeaWifs
Seasonal Range of K490	K490_RG	none	8000	SeaWifs
Mud	MUD	percent	6000	USGS CGS
Benthic Nitrate 1996–2007	NIT_AVG96	μM	40000	GRAMPUS
Benthic Phosphate 1956–1968	PHOS_AVG56	μM	40000	Biochem
Benthic Phosphate 1996–2007	PHOS_AVG96	μM	40000	GRAMPUS
Benthic Salinity 1956–1968	SAL_AVG56	psu	6000	Biochem
Benthic Salinity 1996– 2007	SAL_AVG96	psu	6000	HD NEFSC
Seasonal Range of Benthic Salinity 1996–2007	SAL_RG96	psu	6000	HD NEFSC
Sand	SAND	percent	6000	USGS CGS
Benthic Silicate	SIL_AVG96	μM	6000	GRAMPUS
Slope	SLOPE	degree	650	USGS
Average SST	SST_AVG	degree C	972	SeaWifs
Seasonal Range of SST	SST_RG	degree C	972	SeaWifs
Stratification from 1956–1968	STRAT56	none	2500	HD
Stratification from 1996–2007	STRAT96	none	2500	HD
Summer Stratification from 1956–1968	STRAT_SUM56	none	2500	HD
Summer Stratification from 1996–2007	STRAT_SUM96	None	2500	HD

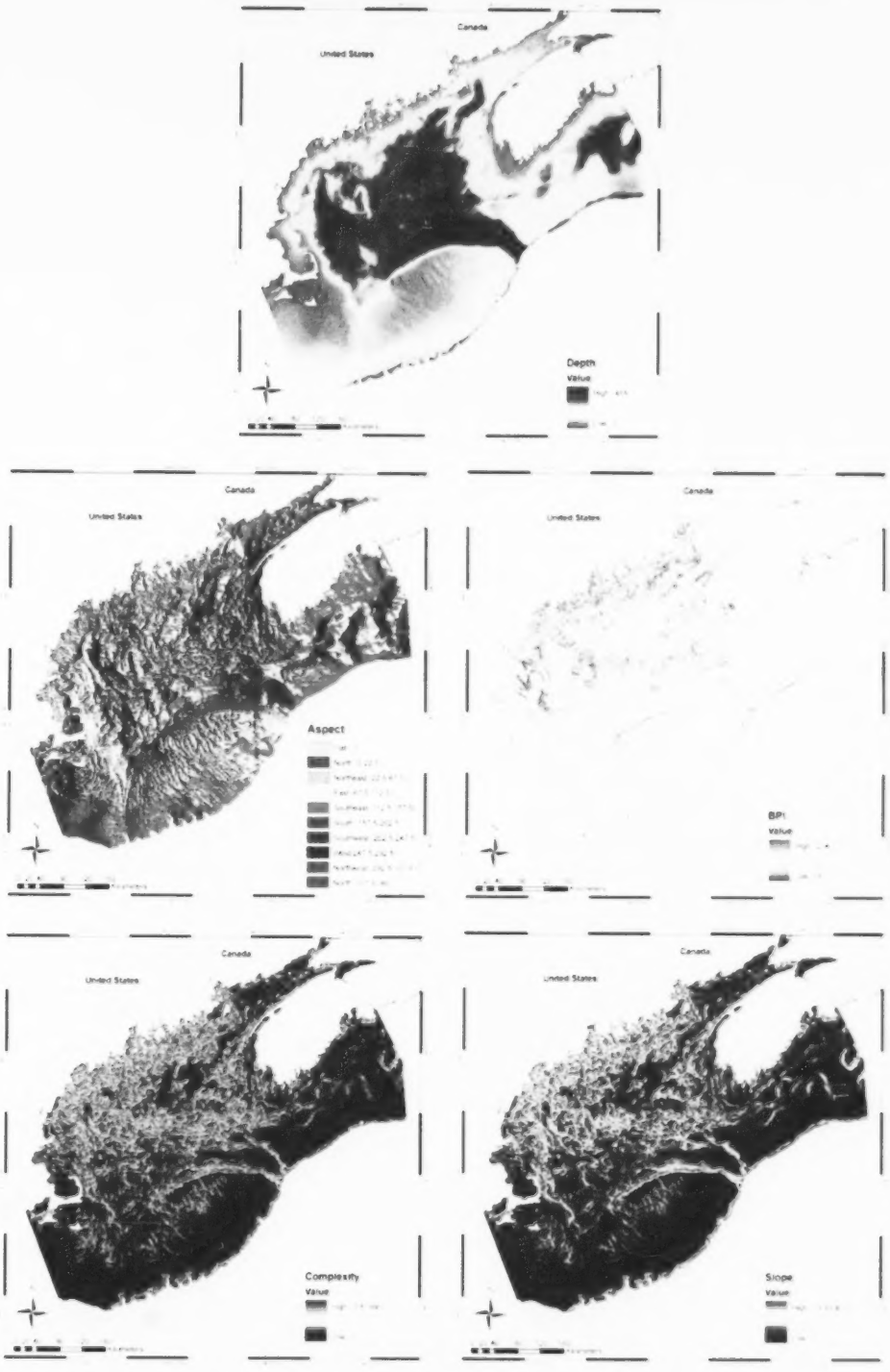


Figure 5. Depth and derivatives created from the USGS 15 arc-second digital elevation model of the GOM.

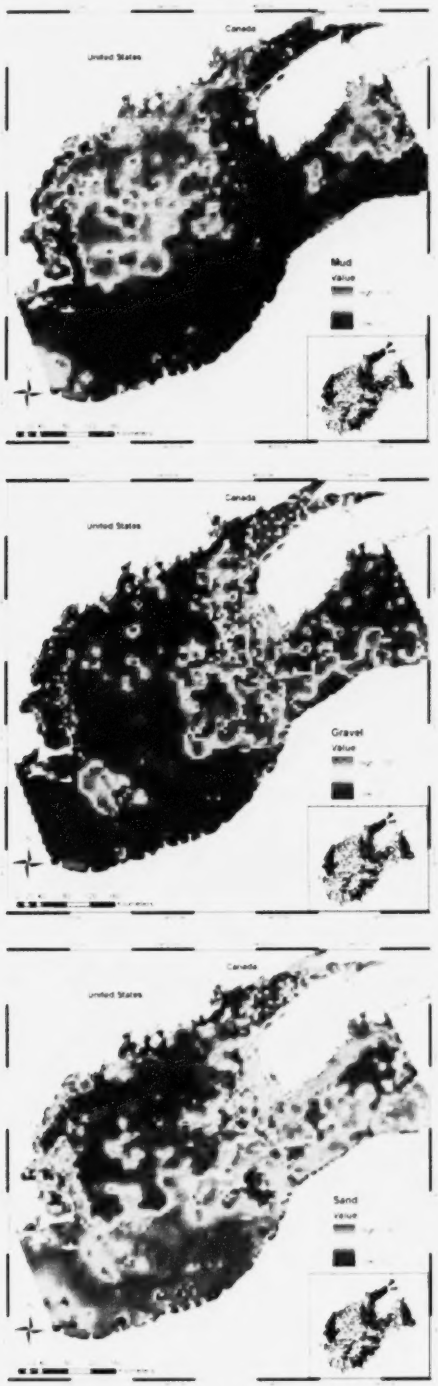


Figure 6. Substrate layers of percent mud, percent gravel, and percent sand(from the USGS and CGS sediment samples).Insert maps in each figure show the density of points used to create the layers.

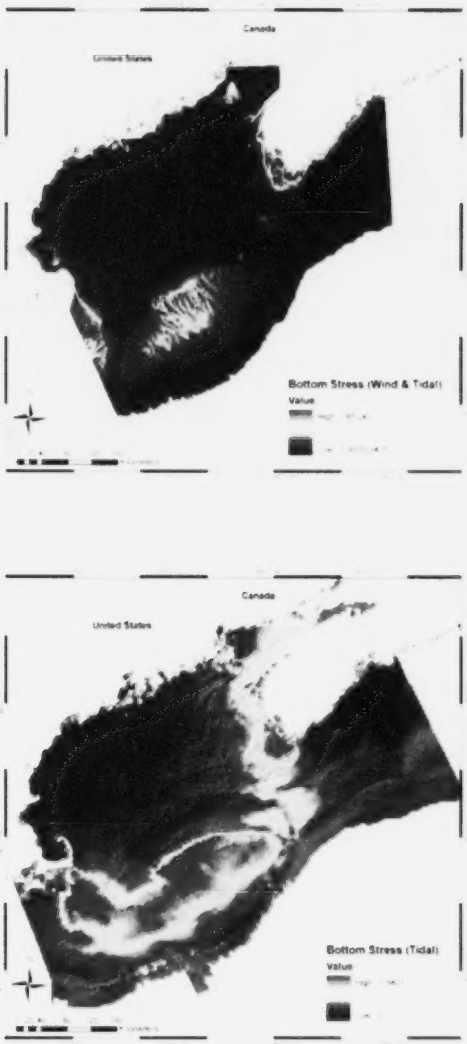


Figure 7.Bottom stress modeled with wind and tidal influence, and bottom stress modeled with only tidal influence. Note that the scales of the two bottom stress layers span over different ranges to emphasize areas of high stress, but at a different magnitude when wind is not included.Original data to create inset maps of bottom stress data density were not available.

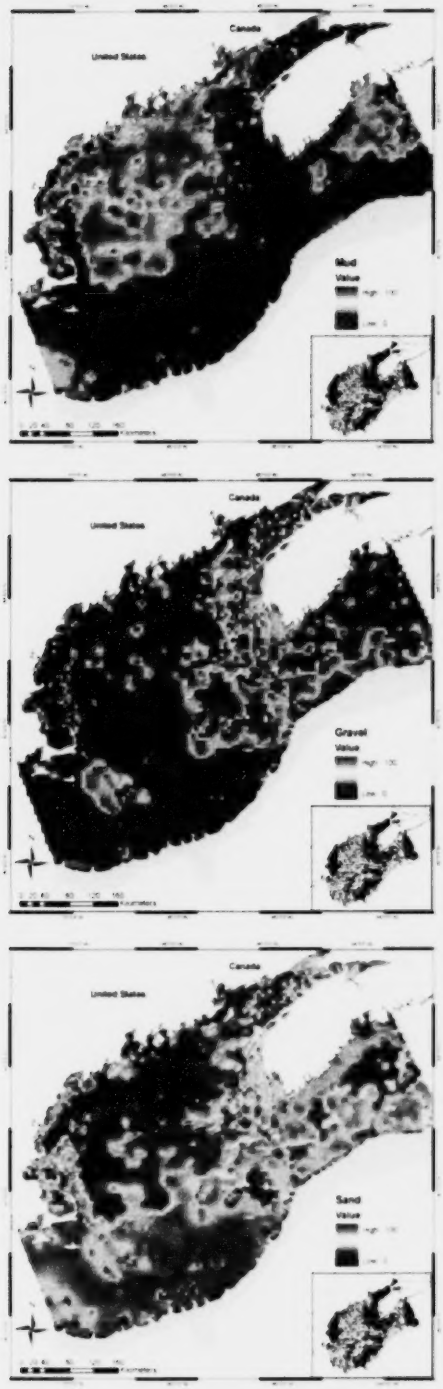


Figure 6. Substrate layers of percent mud, percent gravel, and percent sand(from the USGS and CGS sediment samples).Insert maps in each figure show the density of points used to create the layers.

Figure 7.Bottom stress modeled with wind and tidal influence, and bottom stress modeled with only tidal influence. Note that the scales of the two bottom stress layers span over different ranges to emphasize areas of high stress, but at a different magnitude when wind is not included.Original data to create inset maps of bottom stress data density were not available.

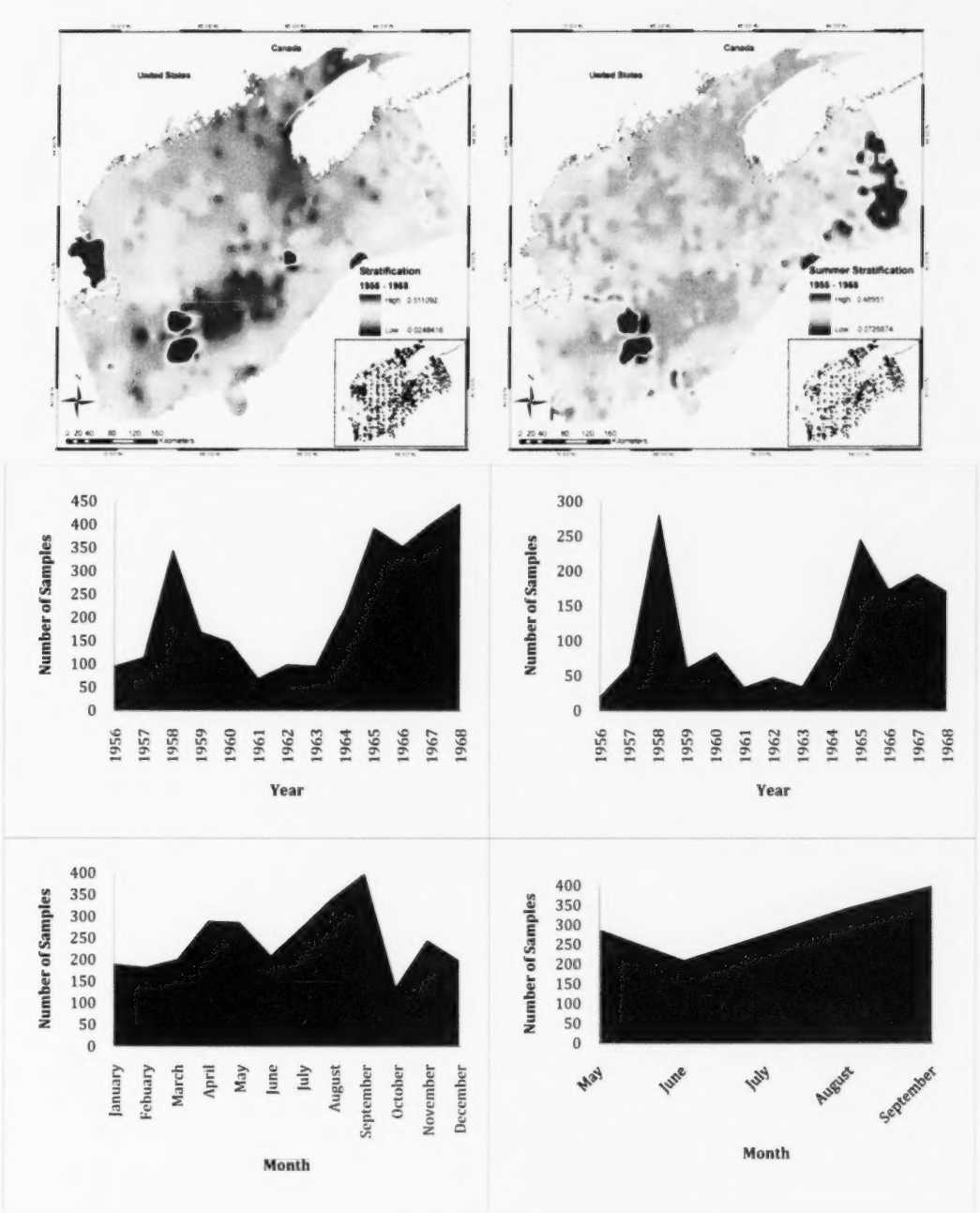


Figure 8. Top: Yearly average and summer stratification for the 1956–1968time-period. Inset maps in each figure show the density of points used to create the layers. Middle: Number of samples per year to create the stratification and summer stratification layers, respectively. Bottom:Number of samples per month to create the stratification and summer stratification layers, respectively.

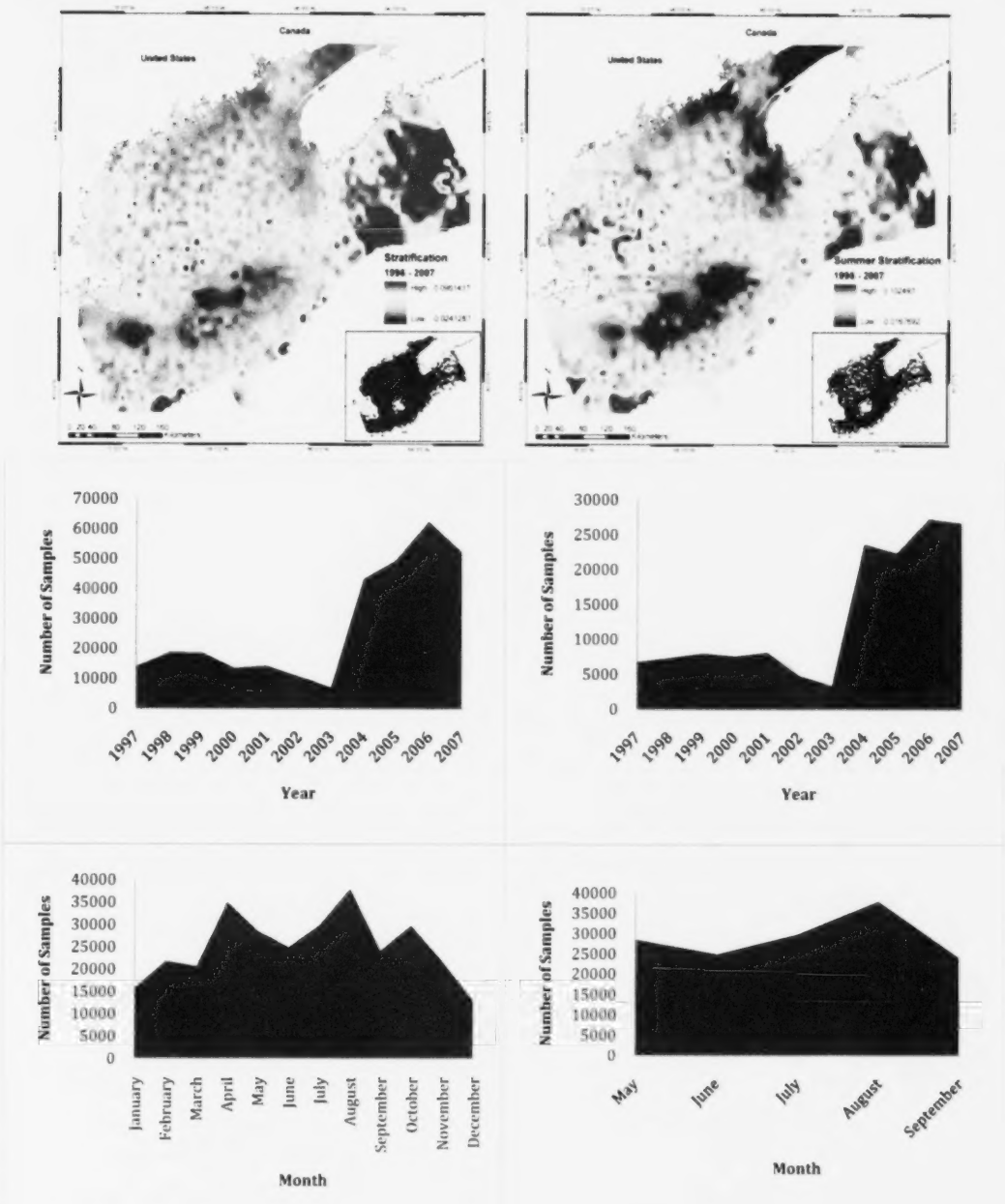


Figure 9. Top: Yearly average and summer stratification for the 1996–2007time-period. Inset maps in each figure show the density of points used to create the layers. Middle: Number of samples per year to create the stratification and summer stratification layers, respectively. Bottom: Number of samples per month to create the stratification and summer stratification layers, respectively.

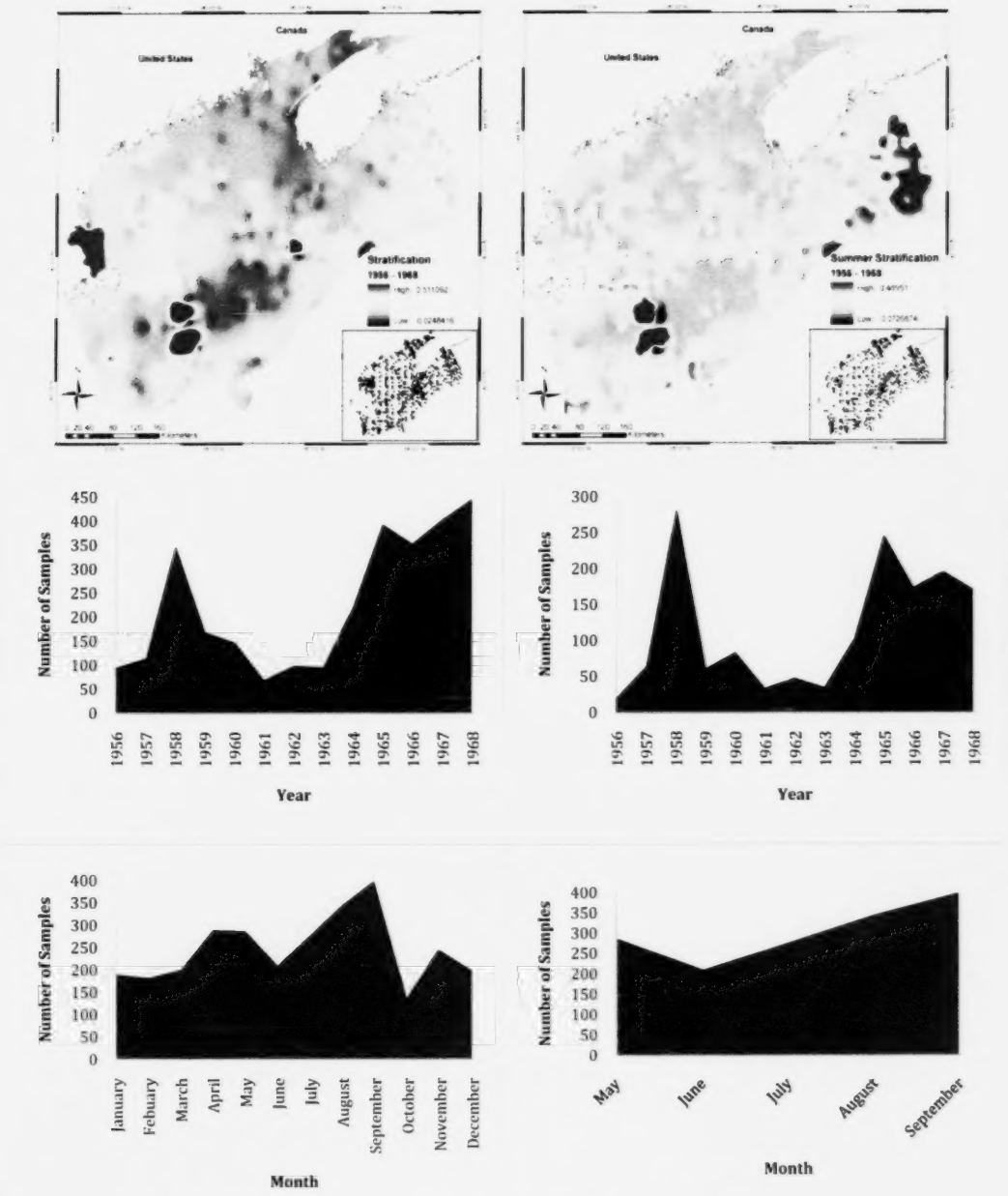


Figure 8. Top: Yearly average and summer stratification for the 1956–1968time-period. Inset maps in each figure show the density of points used to create the layers. Middle: Number of samples per year to create the stratification and summer stratification layers, respectively. Bottom:Number of samples per month to create the stratification and summer stratification layers, respectively.

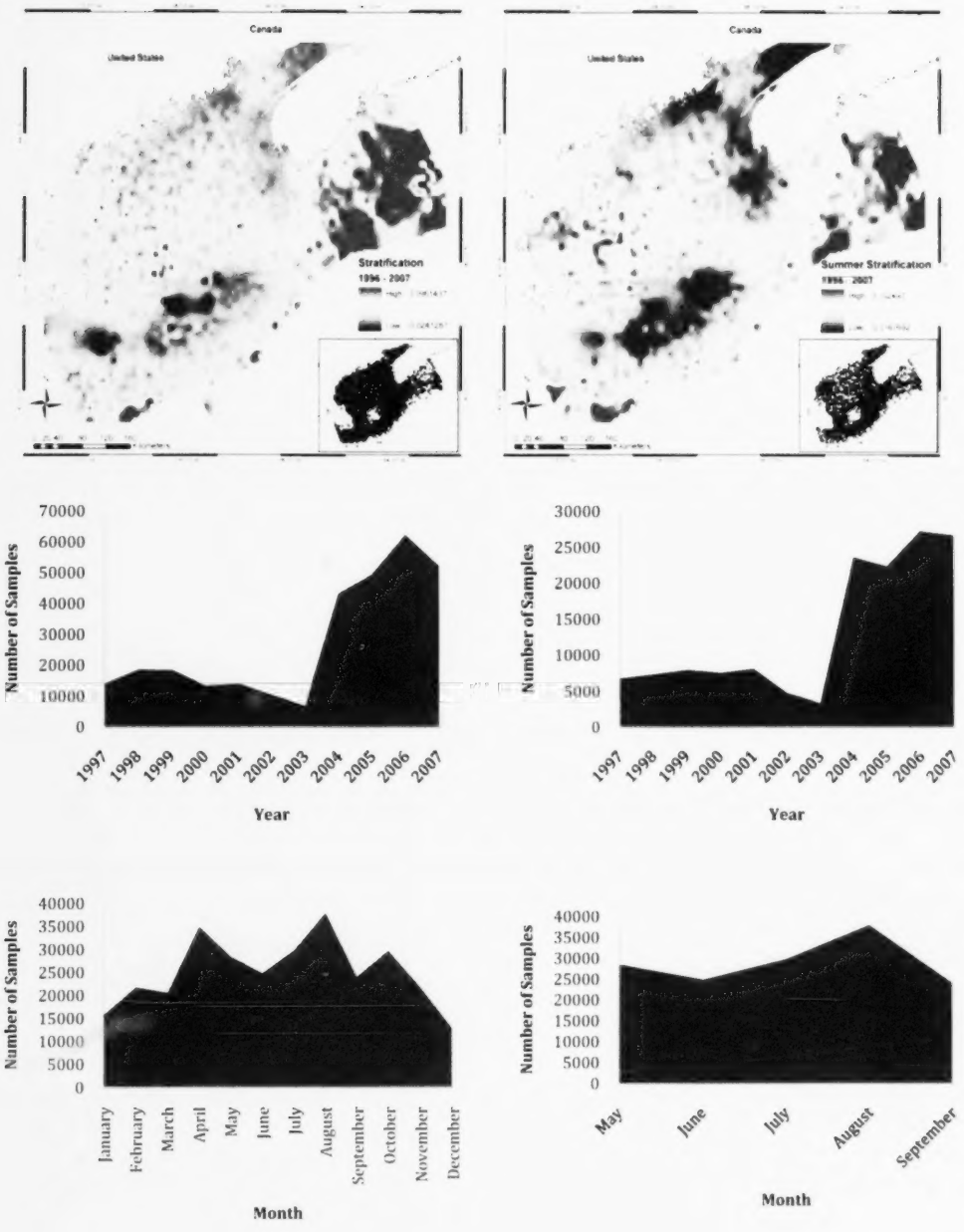


Figure 9. Top: Yearly average and summer stratification for the 1996–2007time-period. Inset maps in each figure show the density of points used to create the layers. Middle: Number of samples per year to create the stratification and summer stratification layers, respectively. Bottom: Number of samples per month to create the stratification and summer stratification layers, respectively.

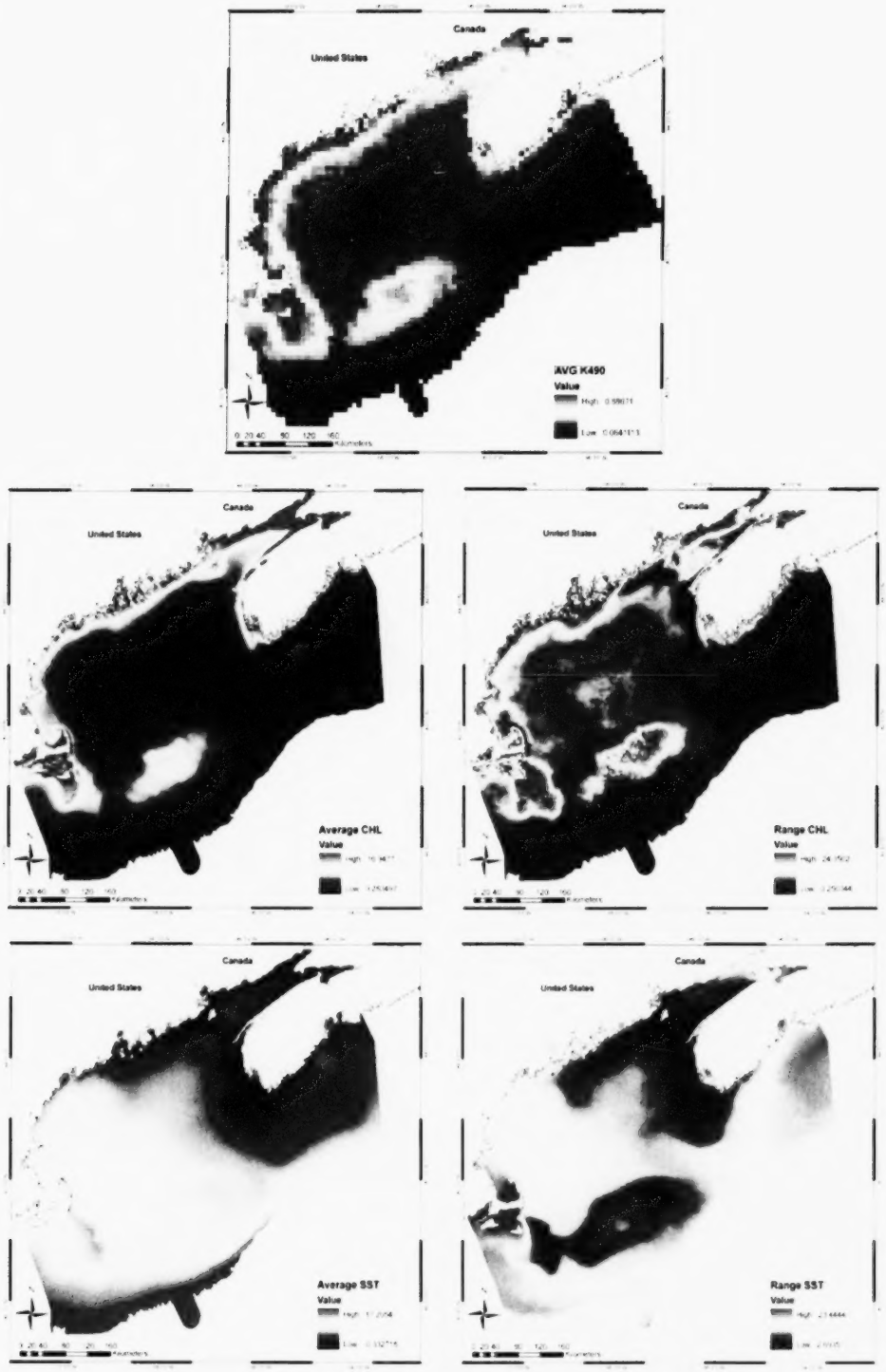


Figure 10. Top: Average SeaWifs K490 from 1997–2008. Middle and bottom: Average and range of chlorophyll (CHL) and sea surface temperature (SST)in the GOM from 1997–2008.

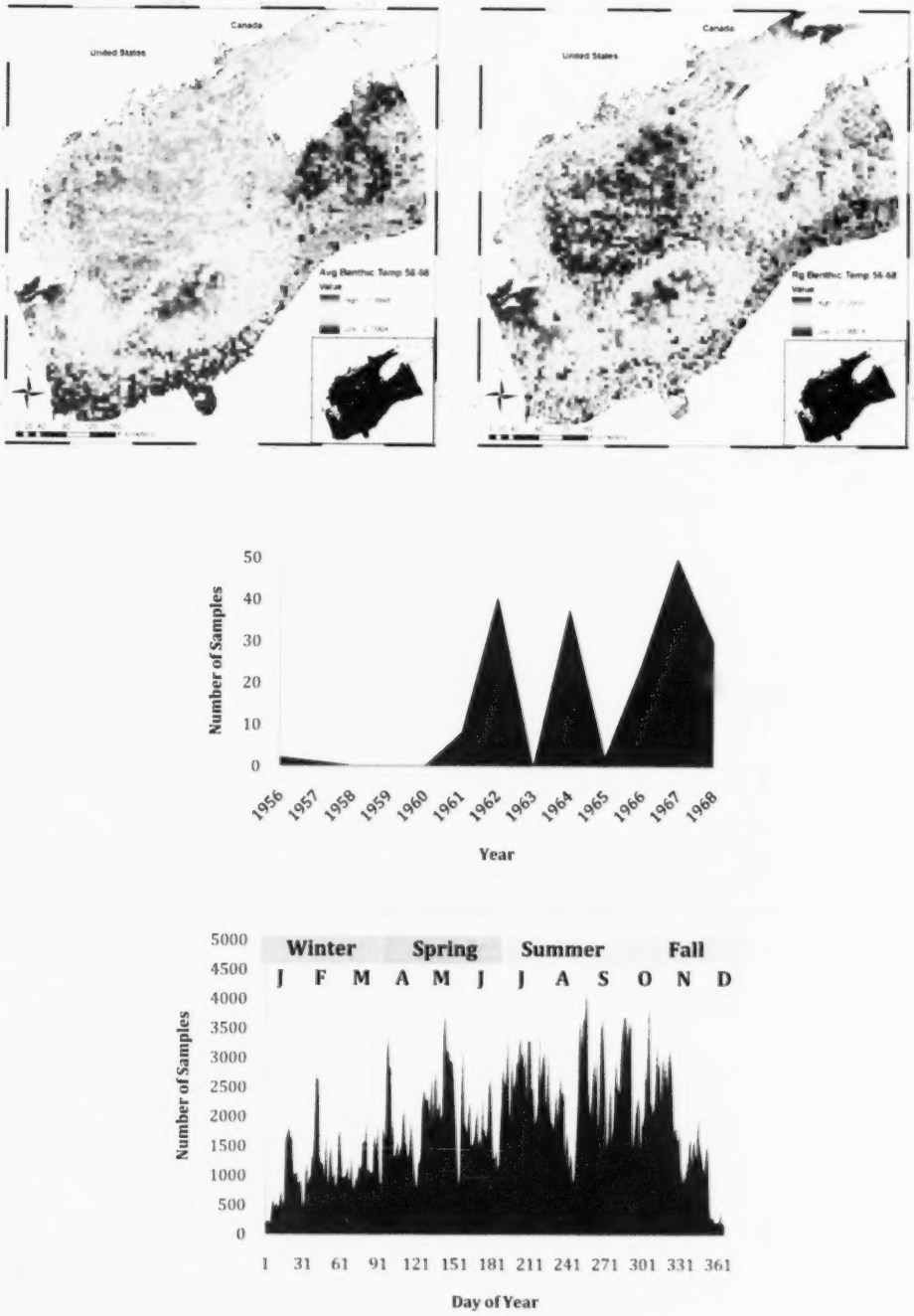


Figure 11. Top: The average and range of benthic temperature from 1956–1968. Inset maps show the density of data used to create that layer. Middle and Bottom: The number of samples used to create the benthic temperature map by year and by day of year, respectively.

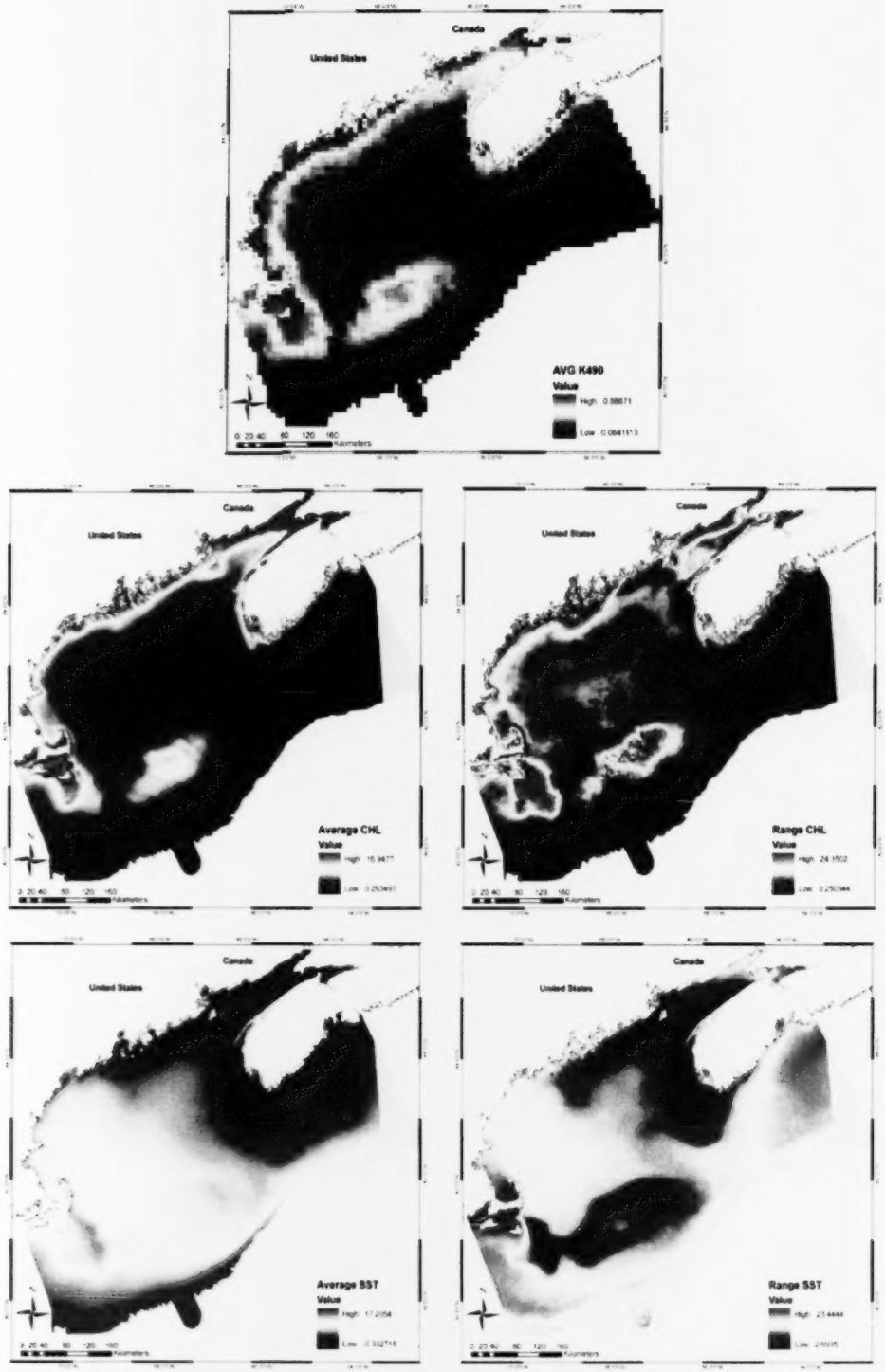


Figure 10. Top: Average SeaWifs K490 from 1997–2008. Middle and bottom: Average and range of chlorophyll (CHL) and sea surface temperature (SST)in the GOM from 1997–2008.

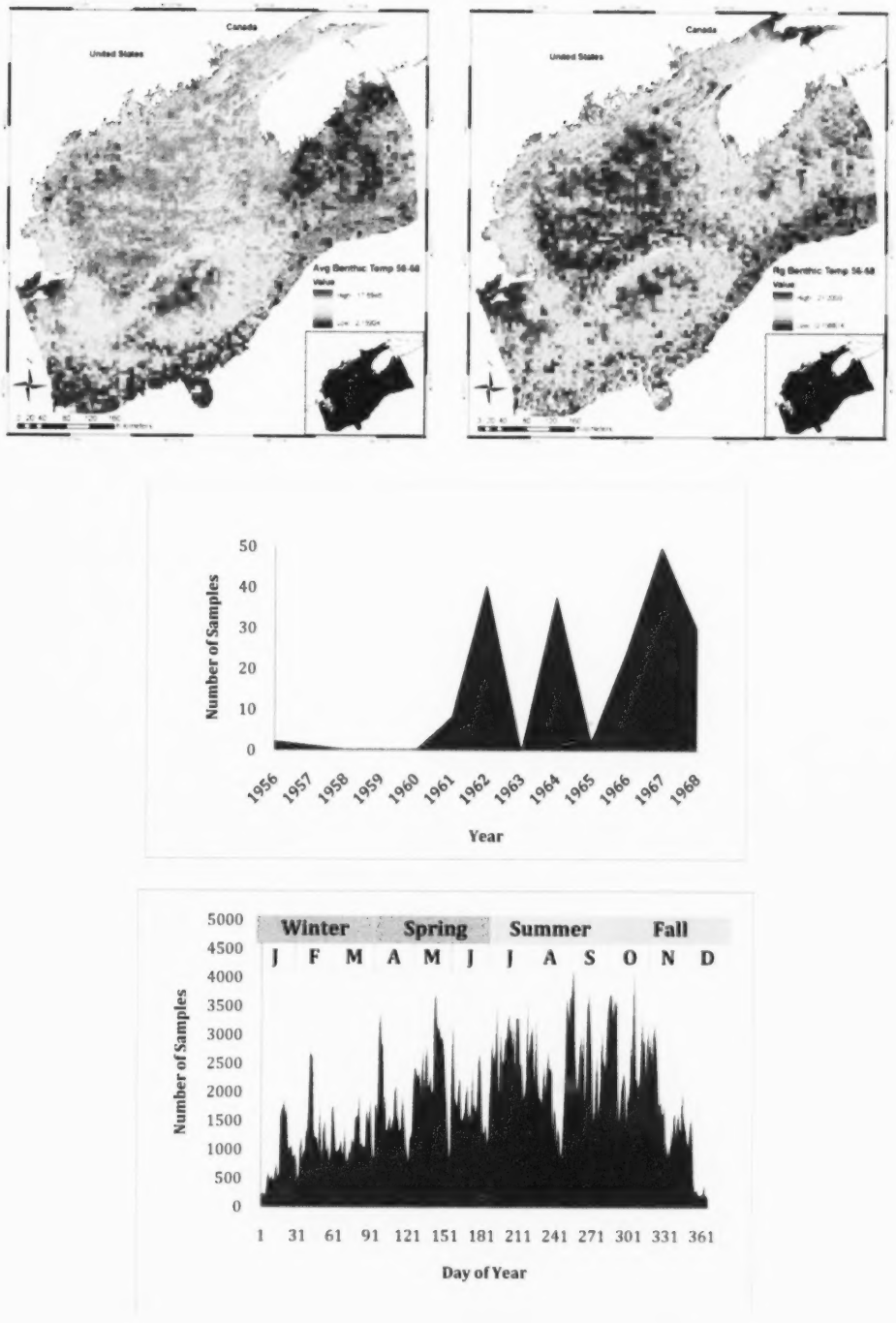


Figure 11. Top: The average and range of benthic temperature from 1956–1968. Inset maps show the density of data used to create that layer. Middle and Bottom: The number of samples used to create the benthic temperature map by year and by day of year, respectively.

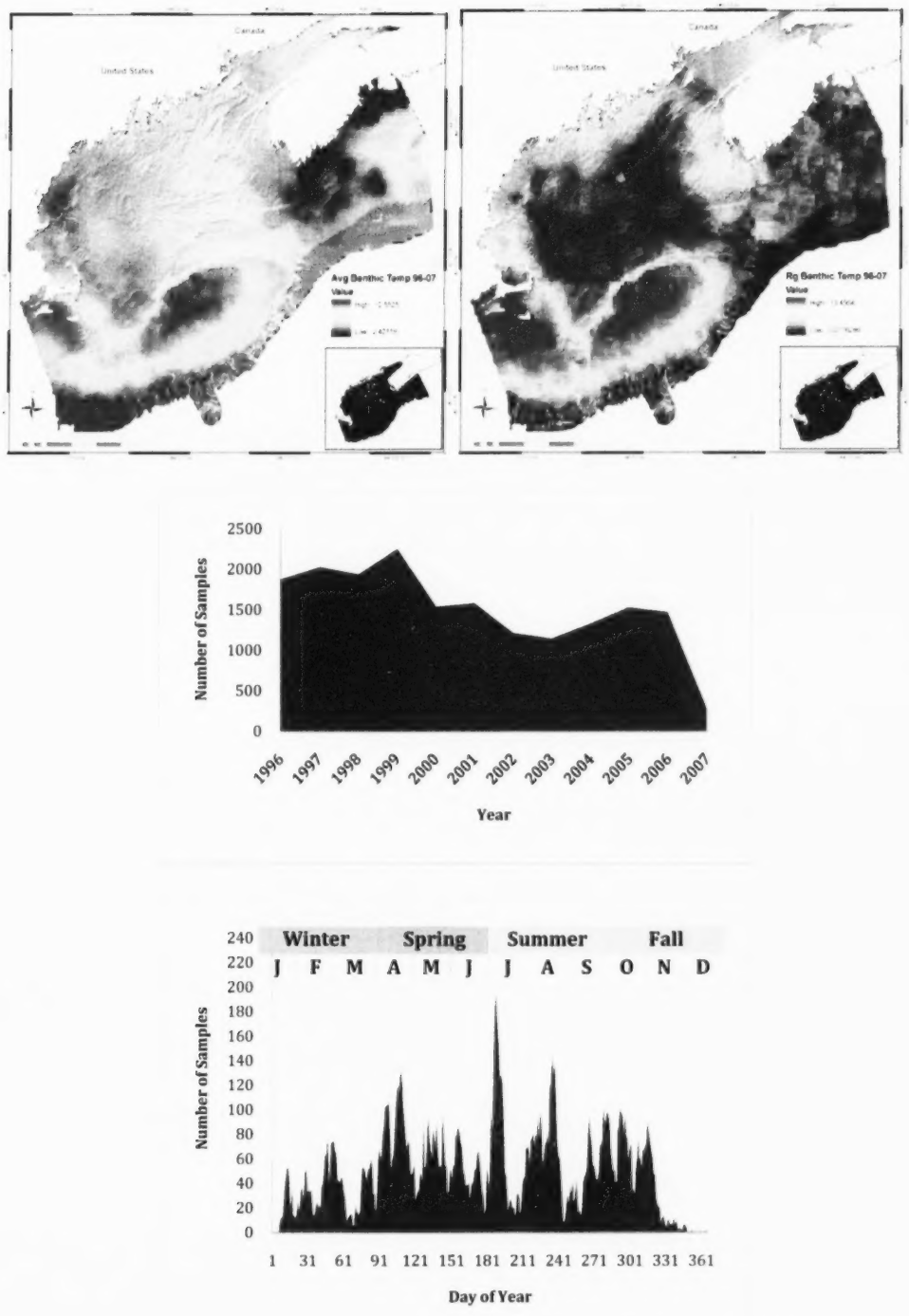


Figure 12. Top: The average and range of benthic temperature from 1996–2007.Inset maps show the distribution of points used to create that layer. Middle and Bottom: The number of samples used to create the benthic temperature map by year and by day of year, respectively.

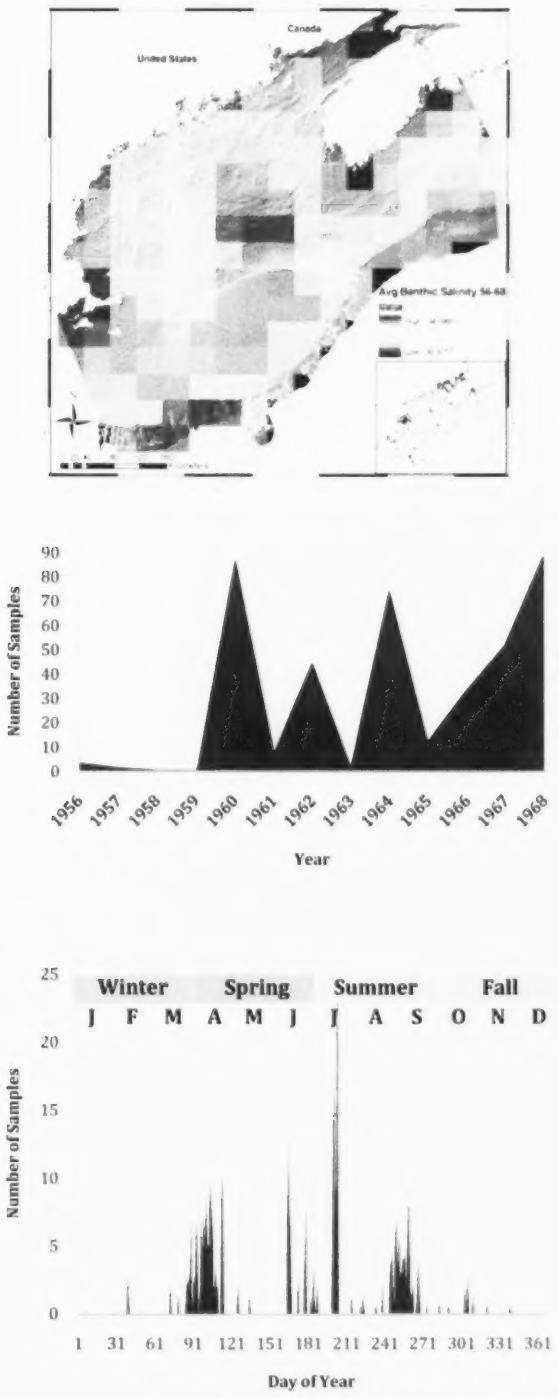


Figure 13. Top: Average benthic salinity from 1956–1968.Inset map shows the distribution of points used to create that layer.Middle and Bottom: The number of samples used to create the benthic salinity map by year and by day of year, respectively.

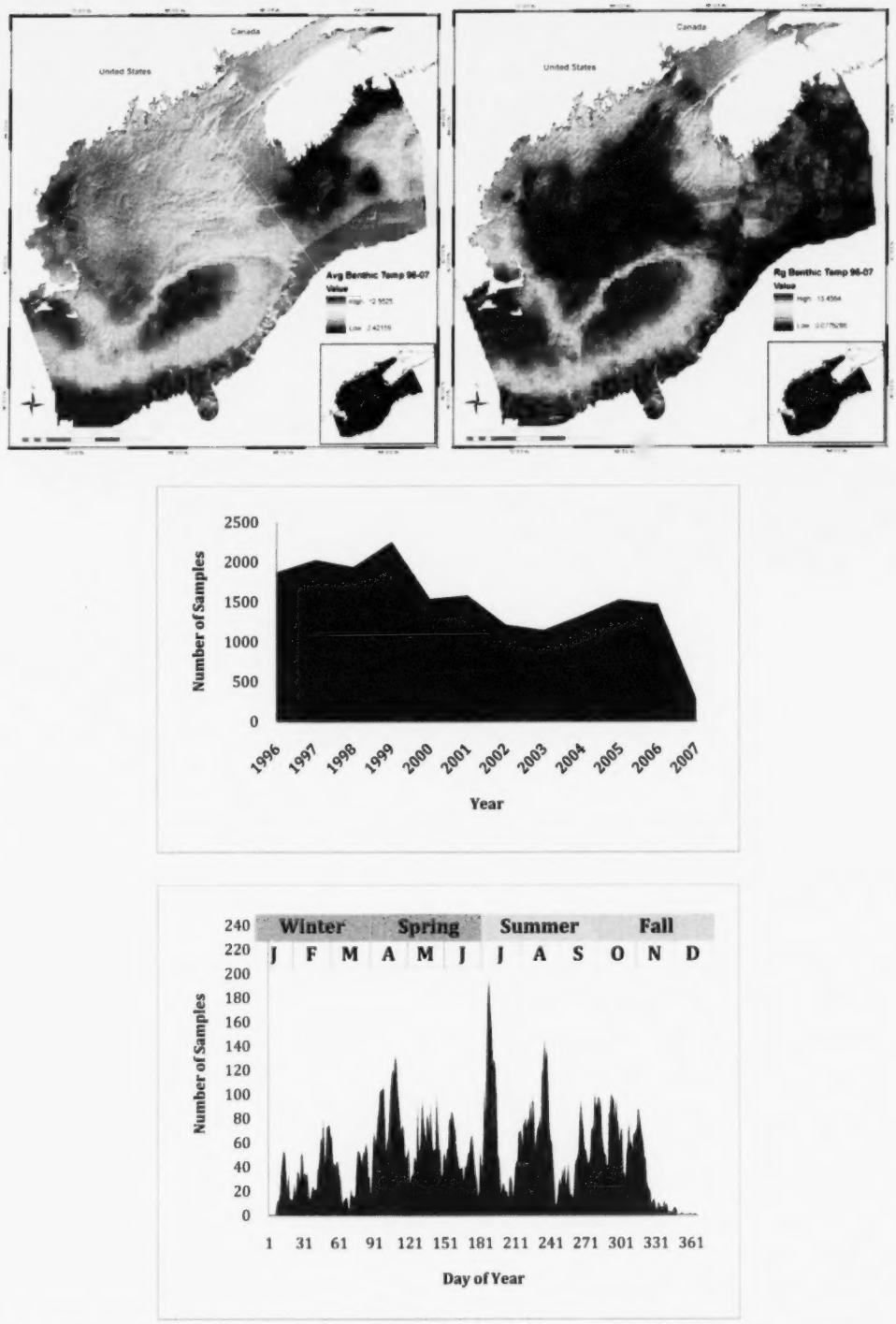


Figure 12. Top: The average and range of benthic temperature from 1996–2007.Inset maps show the distribution of points used to create that layer. Middle and Bottom: The number of samples used to create the benthic temperature map by year and by day of year, respectively.

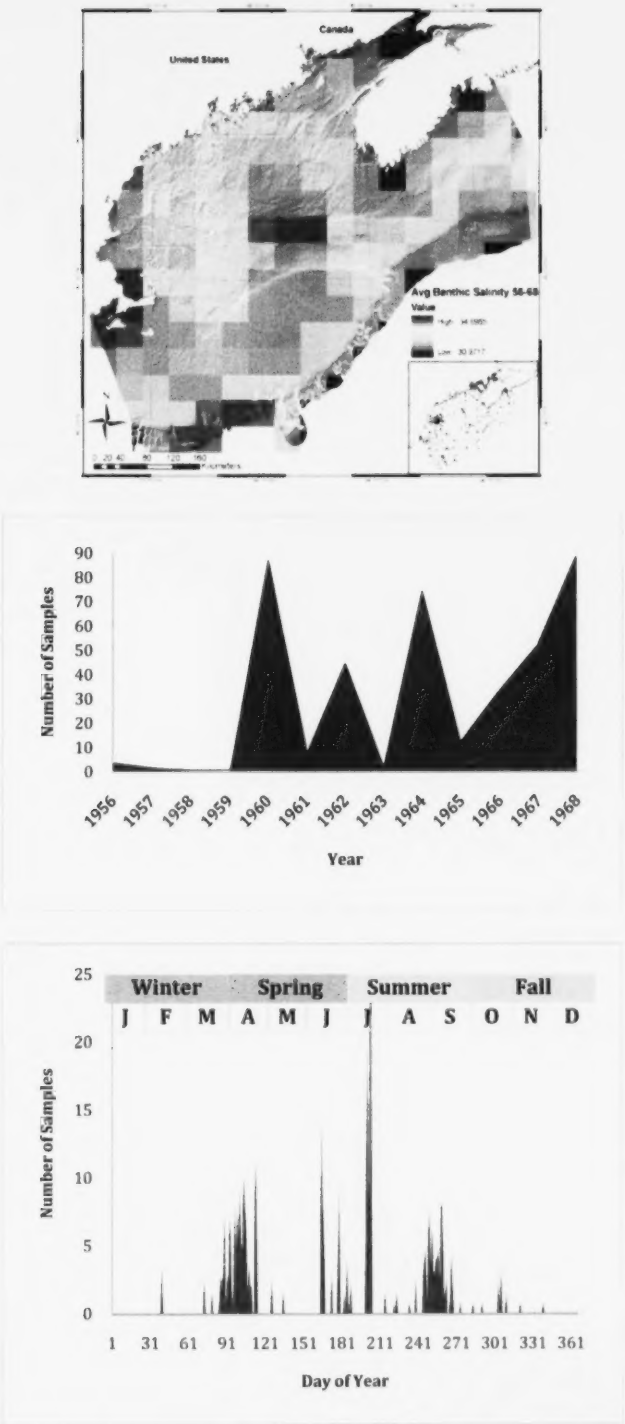


Figure 13. Top: Average benthic salinity from 1956–1968.Inset map shows the distribution of points used to create that layer.Middle and Bottom: The number of samples used to create the benthic salinity map by year and by day of year, respectively.

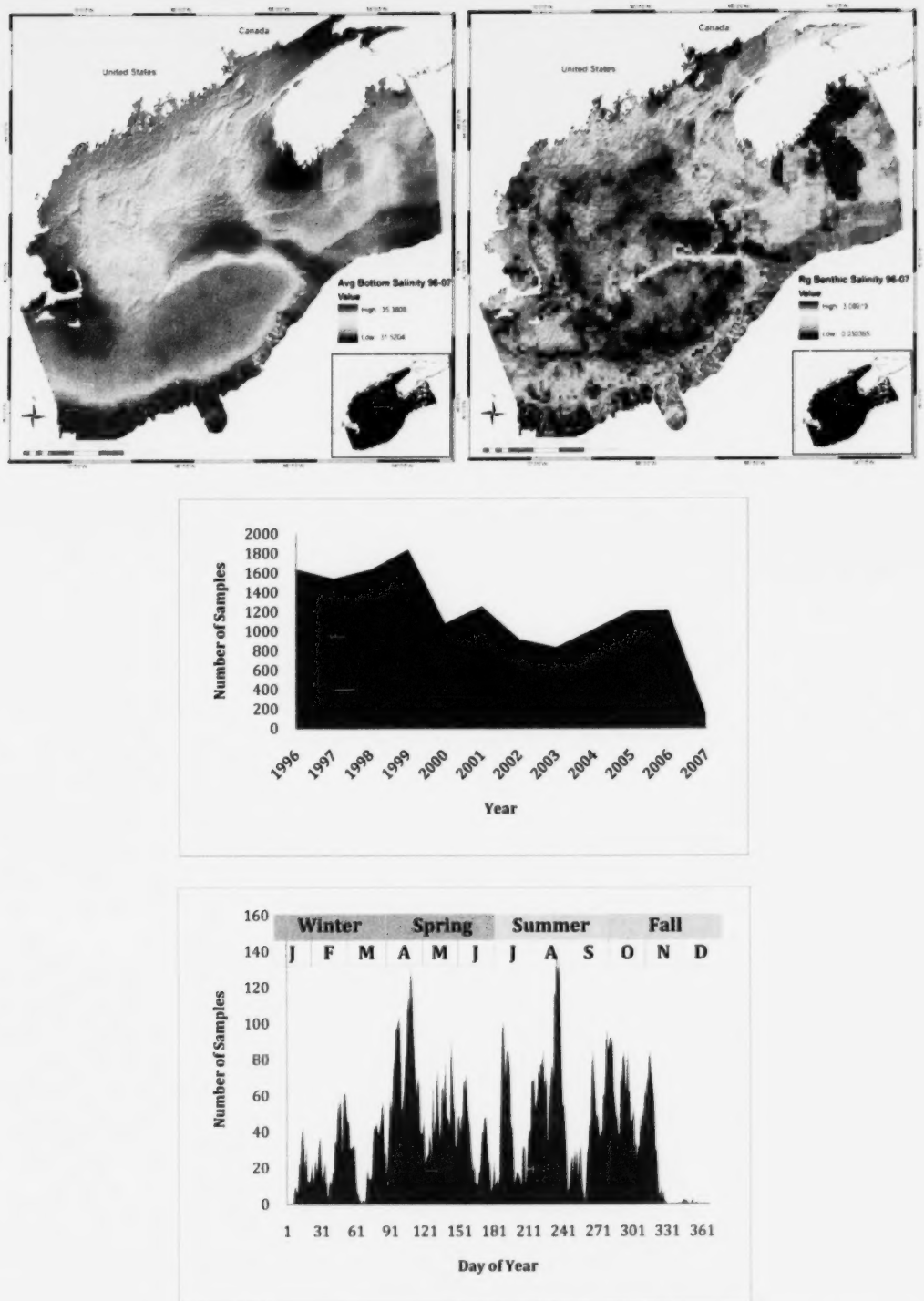
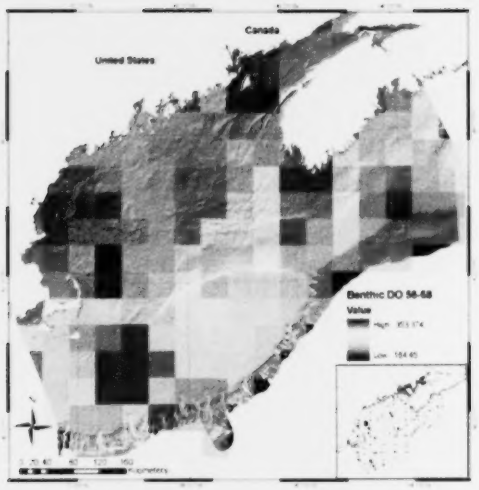
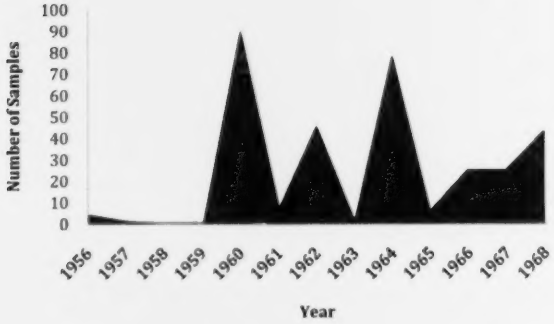


Figure 14.Top: Average and range of benthic salinity from 1996–2007.Inset maps show the density of points used to create that layer.Middle and Bottom: The number of samples used to create the benthic salinity map by year and by day of year, respectively.





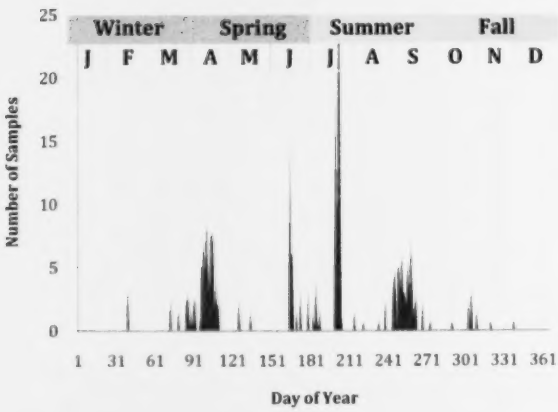
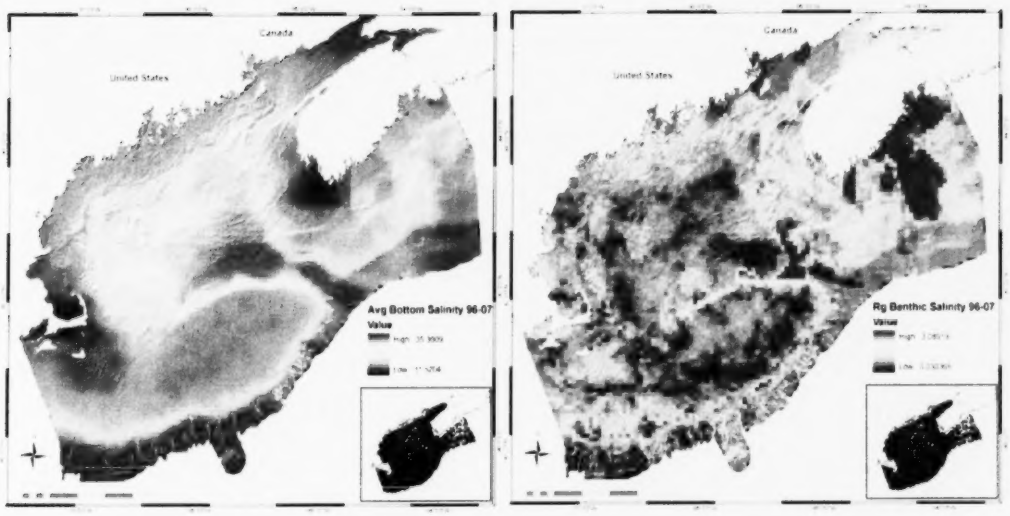
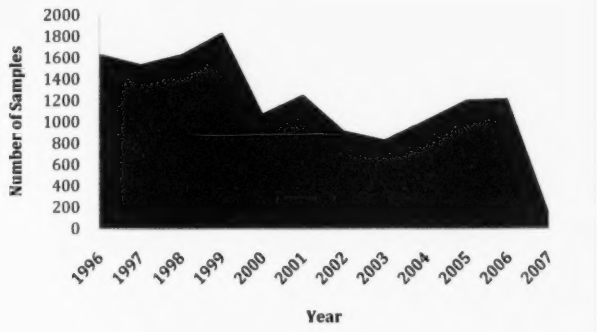


Figure 15. Top: Average benthic dissolved oxygen from 1956–1968.Inset map shows the distribution of points used to create that layer.Middle and Bottom: The number of samples used to create the benthic oxygen map by year and by day of year, respectively.





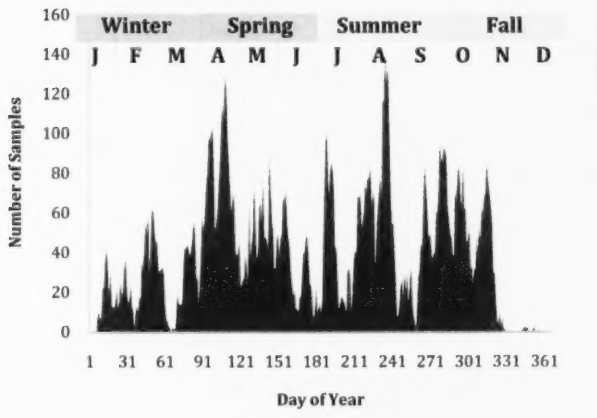


Figure 14.Top: Average and range of benthic salinity from 1996–2007.Inset maps show the density of points used to create that layer.Middle and Bottom: The number of samples used to create the benthic salinity map by year and by day of year, respectively.

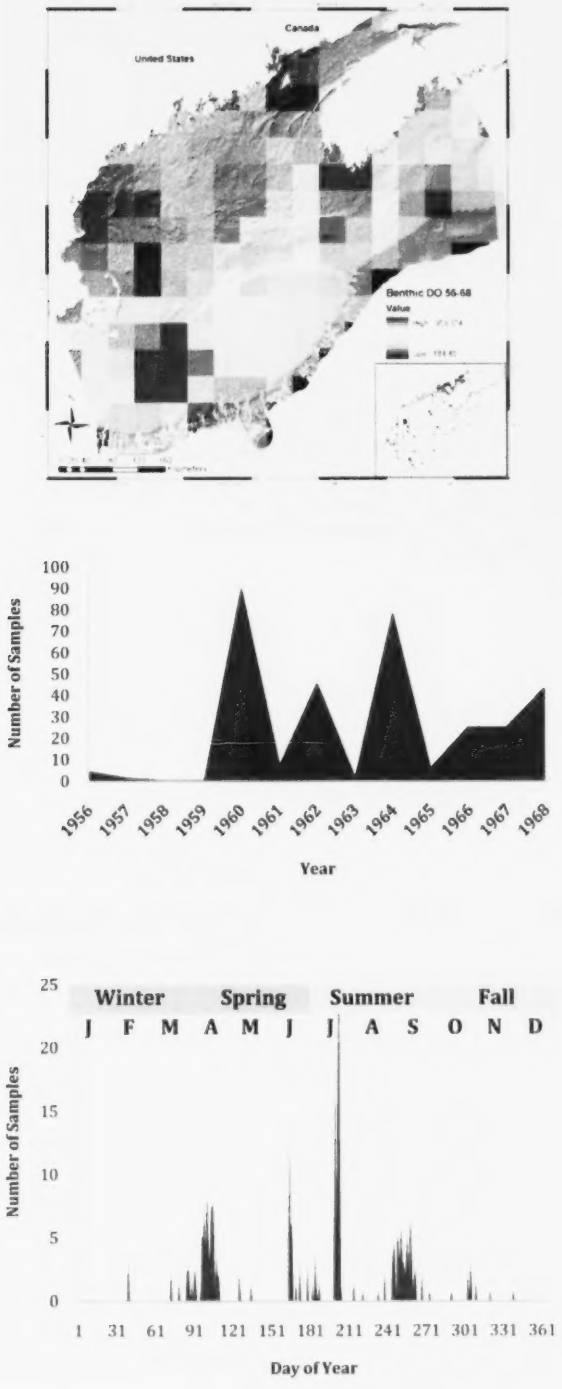


Figure 15. Top: Average benthic dissolved oxygen from 1956–1968.Inset map shows the distribution of points used to create that layer.Middle and Bottom: The number of samples used to create the benthic oxygen map by year and by day of year, respectively.

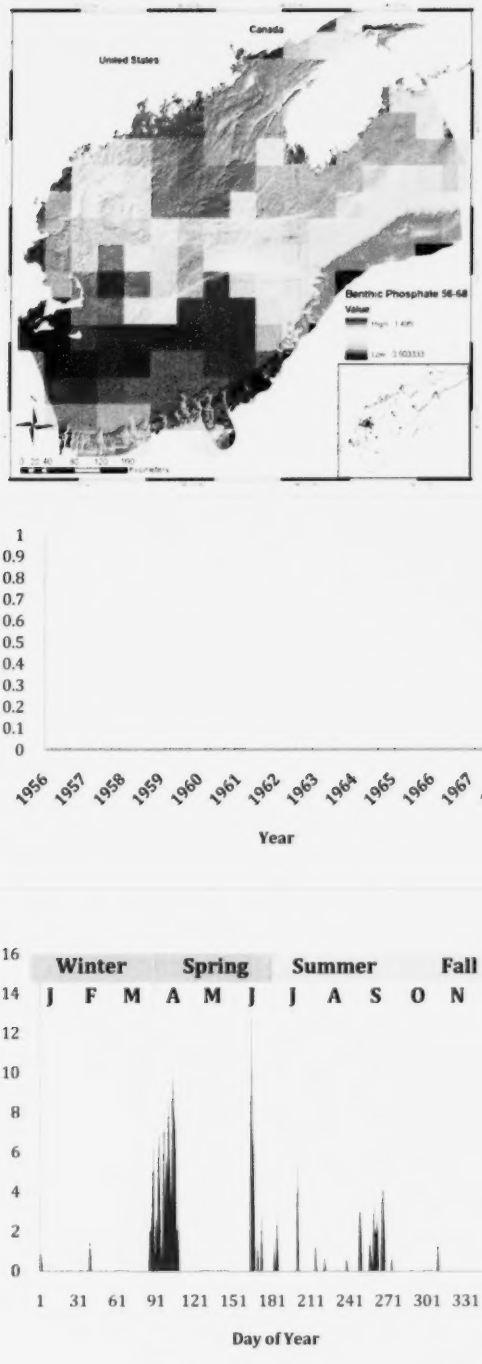


Figure 16. Top: Average benthic phosphate from 1956–1968.Inset map shows the distribution of points used to create that layer.Middle and Bottom: The number of samples used to create the benthic phosphate map by year and by day of year, respectively.

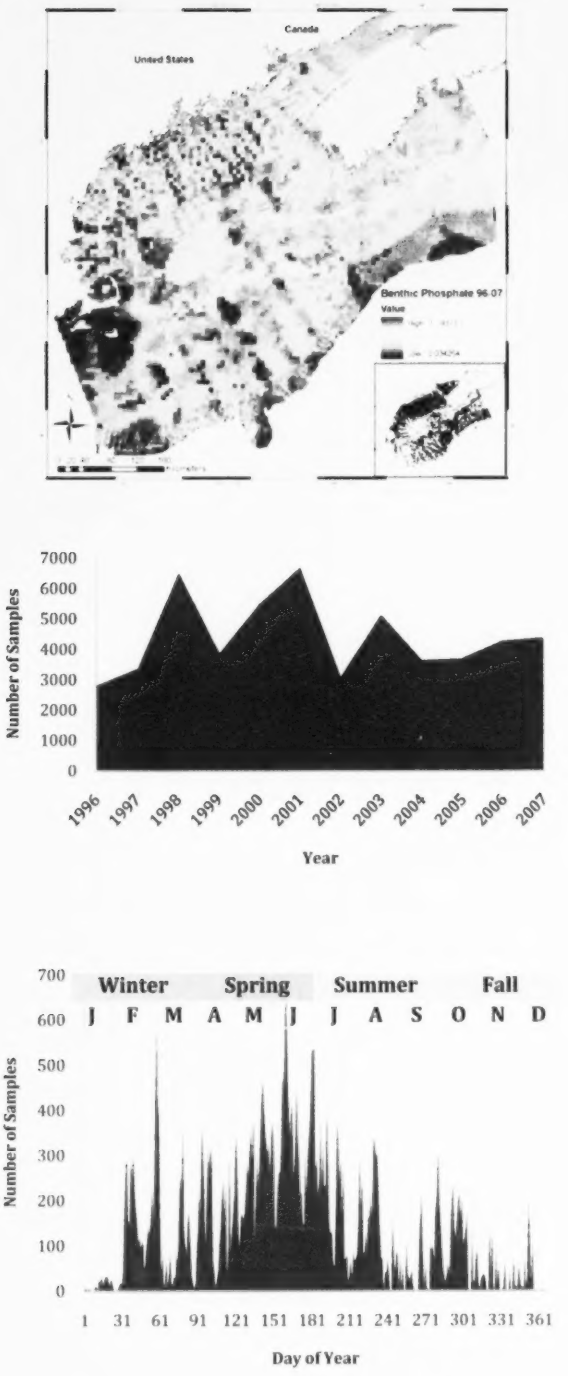


Figure 17.Top:Average benthic phosphate from 1997–2007. Inset map shows the distribution of points used to create that layer.Middle and Bottom: The number of samples used to create the benthic phosphate map by year and by day of year, respectively.

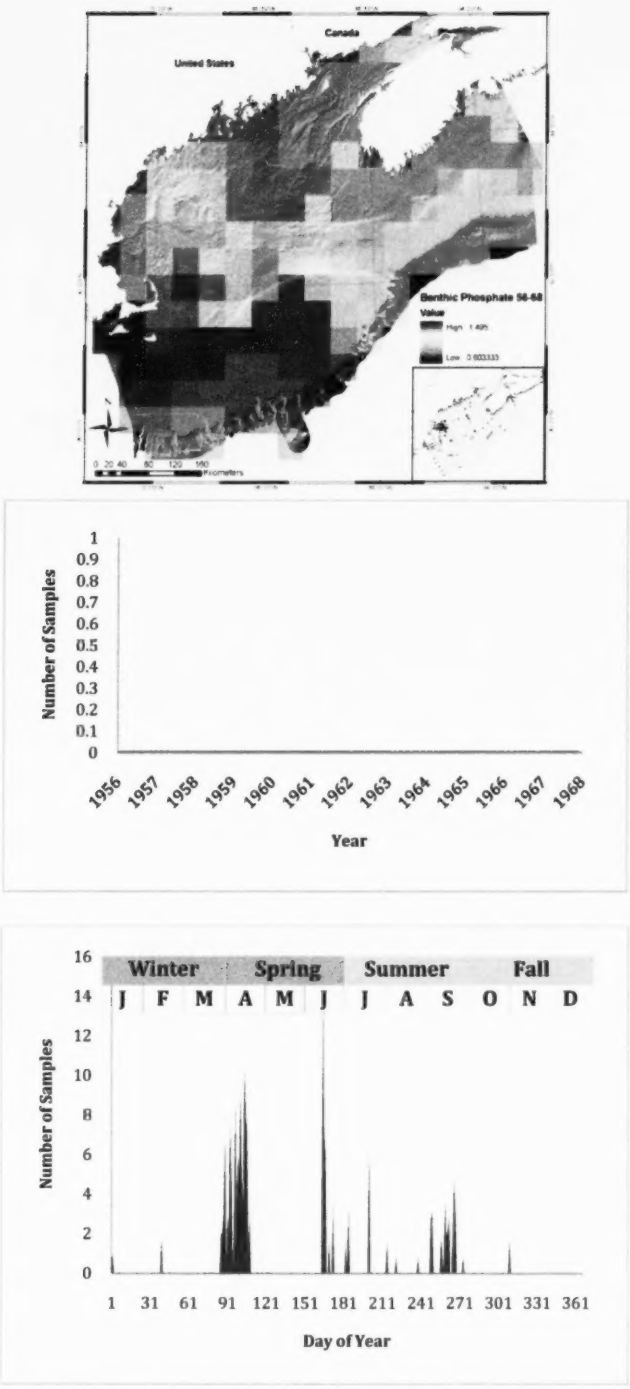


Figure 16. Top: Average benthic phosphate from 1956–1968.Inset map shows the distribution of points used to create that layer.Middle and Bottom: The number of samples used to create the benthic phosphate map by year and by day of year, respectively.

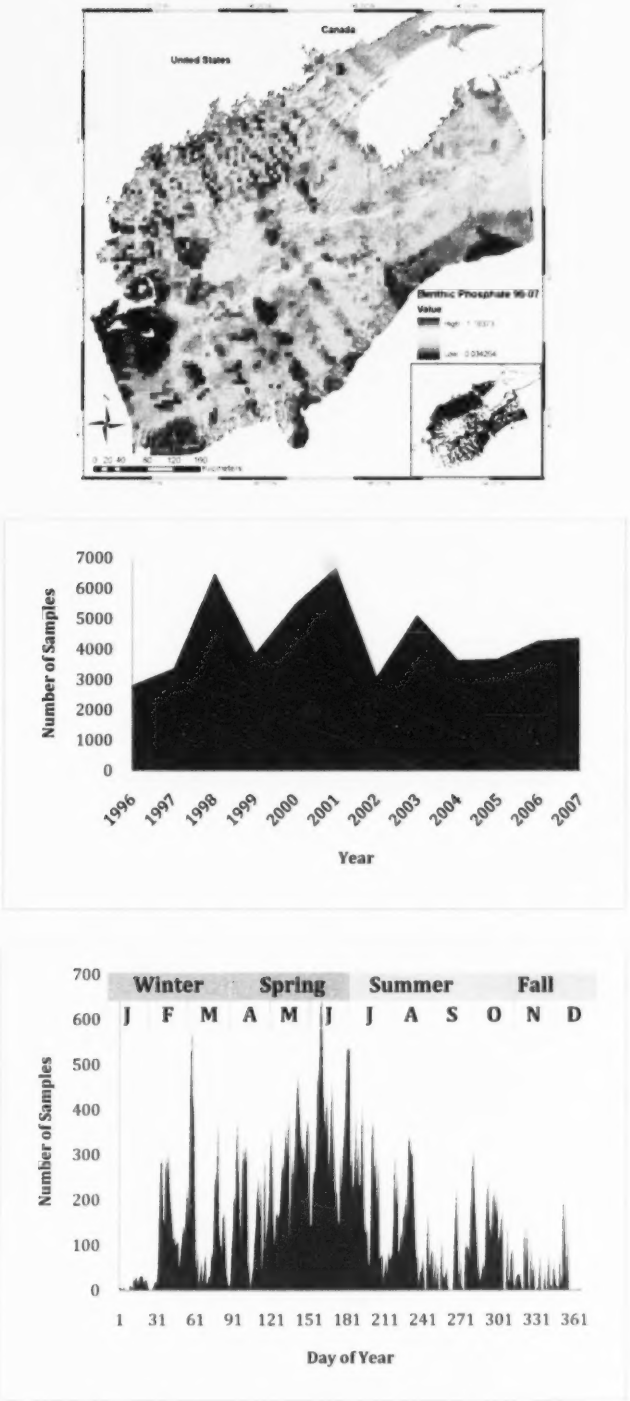
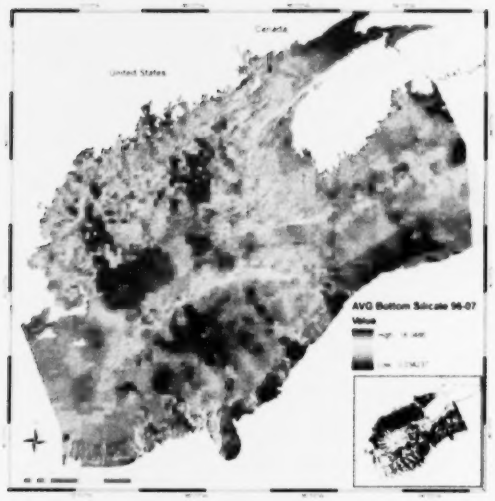
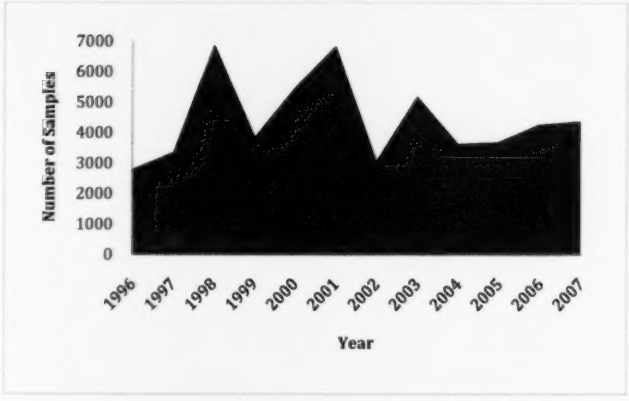


Figure 17.Top:Average benthic phosphate from 1997–2007. Inset map shows the distribution of points used to create that layer.Middle and Bottom: The number of samples used to create the benthic phosphate map by year and by day of year, respectively.





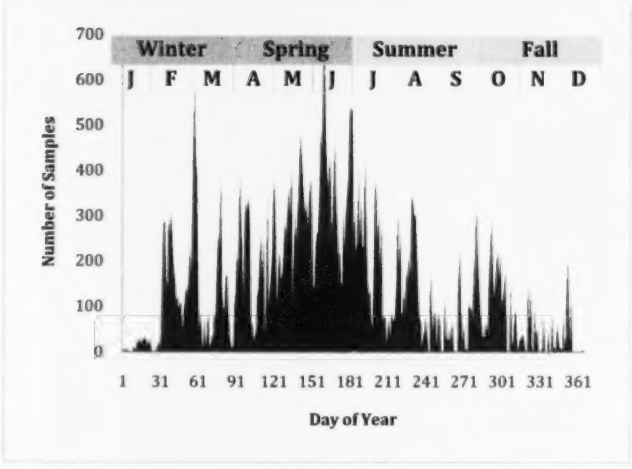


Figure 18. Top:Average benthic silicate from 1996-2007. Insert map shows the distribution of samples used to create that layer. Middle and Bottom: The number of samples used to create the benthic silicate map by year and by day of year, respectively.

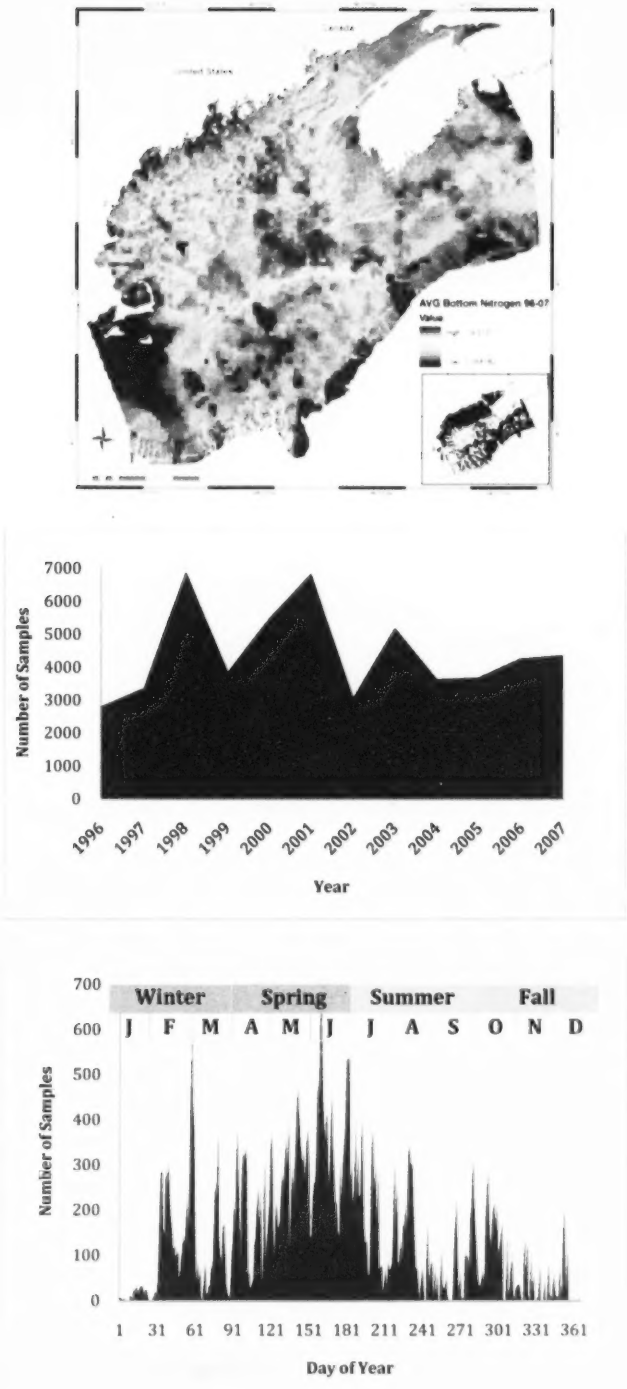


Figure 19. Top: Average benthic nitrate 1996–2007. Inset map shows the distribution of points used to create that layer. Middle and Bottom: The number of samples used to create the benthic nitrate map by year and by day of year, respectively.

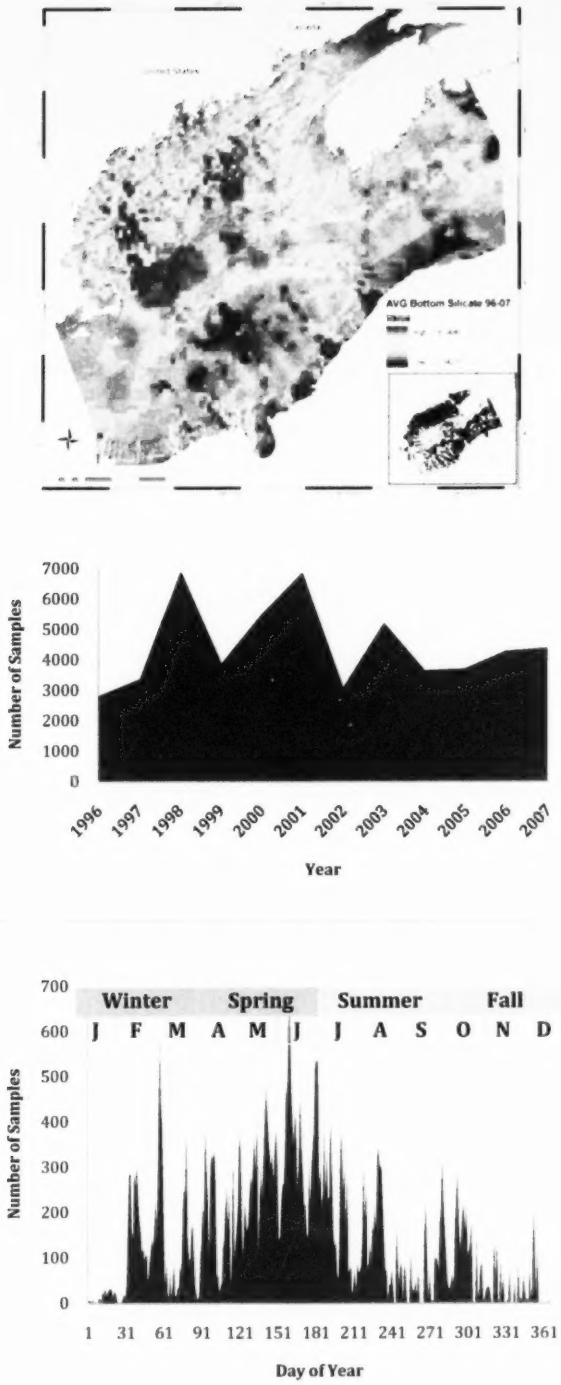


Figure 18. Top:Average benthic silicate from 1996-2007. Insert map shows the distribution of samples used to create that layer. Middle and Bottom: The number of samples used to create the benthic silicate map by year and by day of year, respectively.

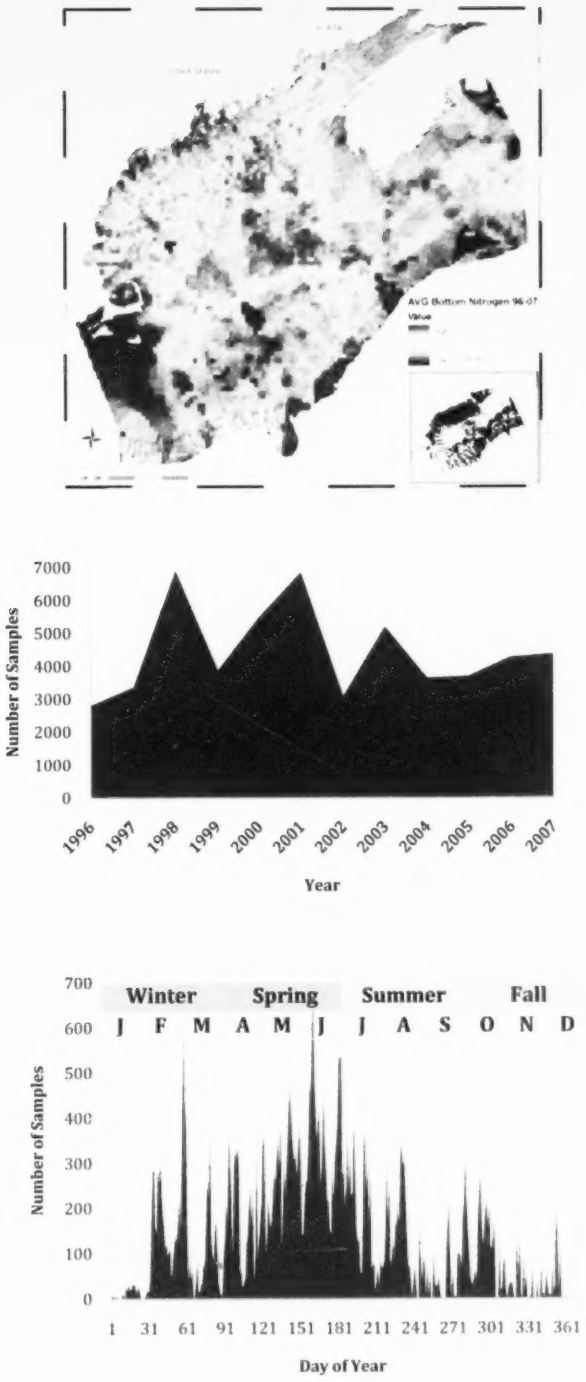


Figure 19. Top: Average benthic nitrate 1996–2007. Inset map shows the distribution of points used to create that layer. Middle and Bottom: The number of samples used to create the benthic nitrate map by year and by day of year, respectively.

DISCUSSION

A comprehensive suite of oceanographic layers for the Gulf of Maine Area has been compiled within a single geodatabase, and in xyz format, using publicly–available U.S. and Canadian oceanographic data sources. The primary driver for this effort was the Census of Marine Life's cross-project synthesis on the role of physical variables in predicting patterns of distribution and diversity in seabed assemblages. Using Gradient Forest (multivariate Random Forests) statistical analyses (Ellis *et al.* 2010 In Prep), the relative importance of different environmental variables in their predictive capacity relative to observed species distributions in three large regional biological survey datasets was evaluated; however these results are outside the scope of this report (see: Pitcher *et al.* (In Prep)).

The preservation of the oceanographic datasets ensures quick and relatively easy access in a common and standard format, thus facilitating their use in future studies. Although large regional spatial physical datasets, similar to the ones presented in this report, have previously been collated for specific research projects on the Scotian Shelf/Western GOM (Day and Roff 2000, Roff *et al.* 2003, Kostylev 2004, Greene *et al.* 2010) they have not generally been made available in a geodatabase or other accessible formats (data from Greene *et al.* 2010 is available, but only includes a few of the variety of layers assembled for our analysis). Currently, many oceanographic physical layers developed for specific research programs are located on local computers or are only publicly available as images in reports. The relative inaccessibility of these types of data is currently a limitation to conducting ecosystem-based research in support of an ecosystem approach to management.

The compiled datasets presented in this report represent what was deemed the best available data for the GOM for the given time periods (1956–1968 and 1996–2007). However, there were data issues when compiling both physical and biological datasets. No single invertebrate dataset, for the contemporary time period, covered the extent of our study area; the GOM. DFO's Ecosystem Surveys and the NMFS groundfish trawl surveys represent a source of invertebrate data; however many species of invertebrates had to be removed from the analyses when it was determined they were not consistently recorded for the 1996–2007 time period (Tremblay *et al.* 2007). DFO is currently attempting to consistently count all invertebrate species during the Ecosystem Surveys, however it will be many years before this dataset represents a useful invertebrate time-series (Clark 2010). Invertebrate datasets from smaller scale sampling efforts, within specific basins or banks, were available for the contemporary time-period, but could not be used because of their limited spatial extent.

The resolution of the oceanographic layers varied significantly (450 m–6 km). For those datasets that were interpolated, the spatial scale at which these layers were produced represent what was considered the highest reasonable resolution given the quantity and distribution of the data. For those data collected through point-base sampling methods (e.g. CTD, biological survey, etc), the seasonal distribution of data was patchy, with most sampling occurring in the summer months. Other data were limited temporally by the timeframe of particular programs (e.g. SeaWiFS; 1997 to present).

Many of the oceanographic layers derived from point-based sampling methods were interpolated using the OAX optimal estimation program. Unfortunately, this program does not provide any information about the inherent variance of the averages produced as layers over the temporal and spatial scales, which would have helped to determine layer suitability for subsequent analyses.

There were a number of data gaps for the oceanographic data. Sediment carbonate was not recorded in either the USGS or Canadian Geological Survey (CGS), samples but would be of interest for future biodiversity analyses since it was found to be influential in determining biodiversity patterns in other regions (Pitcher *et al.* In Prep). Benthic current stress was also an important factor, in the GOM and in the Great Barrier Reef(Pitcher et al. In Prep), however the spatial extent was limited and did not cover the mid and upper Bay of Fundy, Western Scotian Shelf, or Nantucket Shoals. In the CoMLCross-Project analysis, biological samples had to be removed from these areas due to the lack of benthic current stress data, limiting the study from analyzing one of the highest known current stress areas, the Bay of Fundy (Wildish *et al.* 1986).

Some environmental data, such as SST and CHL, were available through DFO but were not accessible in a format that could be converted to a GIS raster. These data therefore had to be acquired from the Satellite Oceanography Laboratory (University of Maine). The issue of GIS compatibility of remote sensing data is being investigated by DFO's BIO Remote Sensing Unit. Furthermore,when available, satellite CHLdata should not be used in nearshore regions and other highly turbid areas like the Bay of Fundy, as these products are known to be a combination of CHL and turbidity. This factor did not end up being important in our analysis, as Bay of Fundy points were removed due to the lack of benthic current stress data available for that region.

The low density of nutrient data from 1956–1968 prevented us from producing layers of nitrate or silicate for this time period. Phosphate data were also limited, but enough observations were available to produce a layer gridded at a very large extent of 40 km. Finally, the original density of the depth samples used to create the USGS's digital elevation model of the GOM was not available.

The compilation of layers described in this report provides a significant amount of oceanographic data for the benthic and sea-surface environment of the GOM for historical and contemporary time periods. As management of marine resources shifts towards an ecosystem approach, the need for spatial data is expected to increase. The storage of the 31 data layers presented here in a geodatabase with associated metadata should facilitate their use in future work.

ACKNOWLEDGEMENTS

This work was supported in part by the Science sector of the Department of Fisheries and Oceans through the Ecosystem Research Initiative. Peter Lawton and Roland Pitcher would also like to acknowledge funding received from the Alfred P. Sloan Foundation, through the Synthesis Program of the International Census of Marine Life, which facilitated workshops and collaboration between authors.

We would like to acknowledge those who have aided in the creation of this technical report and subsequent analyses, including the many scientists that have generously provided us data: Charles Hannah and Brian Petrie for their help with oceanographic layers and the OAX optimal estimation program; HuijieXue for bottom stress data exports from the GOOMOS NowcastForecast System; Don Clark for providing advice on the proper use of the DFO Ecosystem Surveys data; John Tremblay for his technical report and candid advice on using invertebrate data from the DFO Ecosystem Surveys; Richard Karsten and David Wildish for their information on high current stress areas in the GOM; David Mountain for bottom temperature data; Scott Ryan for advice on DFO's nutrient data.

We would also like to thank the two internal DFO reviewers (Glen Harrison and Jerry Black)the library reviewer (Charlotte McAdam) and Fred Page for their thoughtful and constructive comments.

REFERENCES

Azarovitz, T.R. 1981. A brief historical review of the Woods Hole laboratory trawl survey time series. *In* Bottom trawl surveys/releves au chalut de fond. *Edited by* W.G. Doubleday and D. Rivard. pp. 62-67.

Breiman, L. 2001. Random forests. Mach. Learn. **45**(1): 5-32.

Chadwick, E.P., Broodie, W., Colbourne, E., Clark, D., Gascon, D., and Hurlbut, T. 2007. History of annual multi-species trawl surveys on the Atlantic coast of Canada. AZMP Bulletin(6): 68.

Clark, D. 2010. Maritimes region ecosystem surveys. *Edited by* M. Greenlaw, St. Andrews.

Clark, D., and Branton, B. 2007. DFO Maritimes research vessel trawl surveys http://www.iobis.org/OBISWEB/DisplayMetaData.jsp?content=maps_and_metadata/237.html.

Condie, S.A., and Webster, I.T. 1997. The influence of wind stress, temperature and humidity gradients on evaporation from reservoirs. Water Resour. Res. **33**(12): 2812-2822.

Day, J., and Roff, J.C. 2000. Planning for representative marine protected areas: A framework for Canada's oceans. World Wildlife Fund, Toronto, ON. 156 p.

Despres-Patanjo, L.I., Azarovitz, T.R., and Byrne, C.J. 1988. Twenty-five years of fish surveys in the northwest Atlantic: The NMFS northeast fisheries science center bottom trawl survey program. Mar. Fish. Rev. **50**(4): 69-71.

Drozdowski, A., and Hannah, C.G. 2010. Bottom stress maps of the eastern North American continental shelf: Contributions from tidal currents combined with large wave events. Can. Tech. Rep. Hydrogr. Ocean Sci. No. xxxx. In Press.

Ellis, N., Smith, S.J., and Pitcher, C.R. 2010 In Prep. Gradient forests: Calculating importance gradients on physical predictors.

Fisher, W.L., and Rahel, F.J. 2004. Geographic information systems in fisheries American Fisheries Society, Bethesda, Maryland. 275 p.

Fisheries and Oceans Canada. 2006. Biochem: Database of biological and chemical oceanographic data. Version 8 (2005). [Online]. Available from http://www.meds-sdmm/dfo-mpo.gc.ca/biochem/Biocheme.htm.

Frank, K.T. 2004. The Scotian Shelf groundfish trawl survey. AZMP Bulletin **4**: 23-25.

Gee, K. 2007. Marine spatial planning: A theoretical overview [Electronic Version]. Accessed March 24, 2010, Available from http://www.plancoast.eu/files/berlinConference/1stDay/PlanCoast_MSPKiraGee.pdf

Geological Survey of Canada. 2009. Expedition database [Online Database]. Accessed August 2009, Available from http://gdr.nrcan.gc.ca/ed/index_e.php.

Greene, J.K., Anderson, M.G., Odell, J., and Steinberg, N. 2010. The northwest Atlantic marine ecoregional assessment: Species, habitats and ecosystems. Phase one. The Nature Conservancy, Eastern U.S. Division, Boston, MA.

Gregory, D. 2004. Climate: A database of temperature and salinity observations for the northwest Atlantic. Can. Sci. Advis. Sec. Res. Doc. **2004/075**.

Gregory, D., and Narayanan, S. 2003. Biochem: A national archive for marine biology and chemistry data. AZMP Bulletin 3: 11-17 p.

Han, G.Q., and Loder, J.W. 2003. Three-dimensional seasonal-mean circulation and hydrography on the eastern Scotian Shelf. J. Geophys. Res.-Oceans **108**(3136): 21.

Hannah, C.G., Shore, J.A., Loder, J.W., and Naimie, C.E. 2001. Seasonal circulation on the western and central Scotian Shelf. J. Phys. Oceanogr. **31**(2): 591-615.

He, I., Hendry, R., and Boudreau, G. 2003. OAX demonstration and test case. Available from http://www2.mar.dfo-mpo.gc.ca/science/ocean/coastal_hydrodynamics/Oax/oax.html2008].

Helbig, J., and Higdon, J. 2009. A preliminary investigation of two potential stability data products for AZMP. AZMP Bulletin **8**: 40-44.

Incze, L.S., Lawton, P., Ellis, S.L., and Wolff, N.H. 2010. Biodiversity knowledge and its application in the Gulf of Maine area. *In* Life in the world's oceans: Diversity, distribution and abundance. *Edited by* A.D. McIntyre. Oxford: Blackwell Publishing Ltd. In Press. pp. 43-63.

Kostylev, V.E. 2004. Habitat management template for Scotian Shelf habitat mapping. Report to the Maritimes Regional Advisory Process.

Li, J., and Heap, A. 2008. A review of spatial interpolation methods for environmental scientists. Geoscience Australia. 23: 137 p.

Li, M.Z., and Amos, C.L. 2001. SEDTRANS96: The upgraded and better calibrated sediment-transport model for continental shelves. Comput. Geosci. **27**(6): 619-645.

Mountain, D.G., Taylor, M.H., and Bascunan, C. 2004. Revised procedures for calculating regional average water properties for northeast fisheries science cruises. Northeast Fisheries Science Center Reference Document 04-08.

NASA. 2007. Oceancolor forum [Online Database]. Accessed August 2009, Available from http://oceancolor.gsfc.nasa.gov/.

NOAA's National Marine Fisheries Service Northeast Fisheries Science Center. 2005. Northeast fisheries science center bottom trawl survey data [Online Database]. Accessed August 2009, Available from http://www.iobis.org/OBISWEB/DisplayMetaData.jsp?content=maps_and_metadata/1435.html.

Pitcher, R., Lawton, P., Ellis, N., Smith, S.J., Incze, L.S., Wei, C.-L., Greenlaw, M.E., Wolff, N.H., Sameoto, J.A., and Snelgrove, P.V.R. In Prep. The role of physical environmental variables in shaping patterns of biodiversity composition in seabed assemblages.: 20.

Poppe, L.J., Williams, S.J., and Paskevich, V.F. 2005. USGS east-coast sediment analysis: Procedures, database, and GIS data. USGS Open-File Report 2005-1001.

Rebuck, N.D., Townsend, D.W., and Thomas, M.A. 2009. Gulf of Maine region nutrient and hydrographic database [Online Database]. Accessed October 2009, Available from http://grampus.umeoce.maine.edu/nutrients/#Data.

Roff, J.C., Taylor, M.E., and Laughren, J. 2003. Geophysical approaches to the classification, delineation and monitoring of marine habitats and their communities. Aquat. Conserv.: Mar. Freshwat. Ecosyst. **13**(1): 77-90.

Roworth, E., and Signell, R.P. 1998. Construction of a digital bathymetry for the Gulf of Maine [Online Publication]. Accessed August 2009, Available from http://pubs.usgs.gov/of/1998/of98-801/.

Shackell, N., and Frank, K.T. 2003. Marine fish diversity on the Scotian Shelf, Canada. Aquat. Conserv.: Mar. Freshwat. Ecosyst. **13**: 305-321.

Swail, V.R., and Cox, A.T. 2000. On the use of NCEP-NCAR reanalysis surface marine wind fields for a long-term north Atlantic wave hindcast. J. Atmos. Ocean. Technol. **17**(4): 532-545.

Theroux, R.B., and Wigley, R.L. 1998. Quantitative composition and distribution of the macrobenthic invertebrate fauna of the continental shelf ecosystem of the northeastern United States. NOAA Technical Report NMFS 140: 240 p.

Tremblay, M.J., Black, G.A.P., and Branton, R.M. 2007. The distribution of common decapod crustaceans and other invertebrates recorded in annual ecosystem surveys of the Scotian Shelf 1999 - 2006, Dartmouth. Can. Tech. Rep. Fish. Aquat. Sci. 2762: 78 p.

Wildish, D.J., Peer, D.L., and Greenberg, D.A. 1986. Benthic macrofaunal production in the Bay of Fundy and the possible effects of a tidal power barrage at economy point - Cape Tenny. Can. J. Fish. Aquat. Sci. **43**(12): 2410-2417.

Wood, L.J., and Dragicevic, S. 2007. GIS-based multicriteria evaluation and fuzzy sets to identify priority sites for marine protection. Biodiversity Conserv. **16**(9): 2539-2558.

Xue, H., Shi, L., Cousins, S., and Pettigrew, N.R. 2005. The GOMOOS Nowcast Forecast System,. Cont. Shelf. Res. **25**(17): 2122-2146.